International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 10, No. 7 (2017), pp.93-104
http//dx.doi.org/10.14257/ijsip.2017.10.7.09

High Frequency Bandwidth Extension for Audio Codec in Mobile Surveillance Device

Bo Hang*, Yi Wang and Changqing Kang

College of Mathematics and Computer Science, Hubei University of Arts and Science, Xiangyang, Hubei, 441053, China
E-mail:bohang@163.com

Abstract

To meet the requirements of low coding bit-rate and low complexity for audio coding in mobile surveillance device, in this paper we proposed an audio Bandwidth Extension (BWE) algorithm based on a hybrid prediction model including the intra-frame prediction, inter-frame prediction and white-noise prediction. In the algorithm, we used four different predicting modes, including translation and fold of low-frequency signals, high-frequency signal in previous frame, and white noise, to reconstruct high-frequency signals for various types of audio signals. When a high frequency frame is encoded, Signal-Noise-Ratios (SNR) of all modes are calculated and compared with each other. The prediction pattern with highest SNR is selected as the most accurate prediction pattern to encoding the high frequency signal. And two indicator bits are used indicate the encoding mode. When the compressed high frequency frame is decoded, the specific mode is selected based on the indicator bits. The testing results show that in the same bit-rate, the quality of the proposed BWE is better than the BWE method in Audio Video Standard (AVS) Part 10, and the computational complexity reduced evidently. These advantages help the proposed method to meet the requirements of mobile surveillance device for audio codec.

Keywords*: Bandwidth Extension, Mobile, Audio codec, Surveillance*

1. Introduction

According to the research on human auditory features, human ears are more sensitive to low frequency signals than high frequency signals in the auditory area. And the main information of speech signals is embodied in low frequency signals below 4 kHz. Consequently, the sampling rates of all the early speech encoders are never higher than 8 kHz, and all the generated signals are narrow-band speech signals. But because the high-frequency part above 4 kHz is important in distinguishing unvoiced speech, such as /f/ and /s/, losing high frequency signal will decline the intelligibility of speech, and the naturalness as well [1]. In order to be compatible with existing narrow-band speech signals and find the losing high frequency signals, high-frequency bandwidth extension (BWE) is applied to extend the narrow band speech signals to wide band. The method is utilizing the features of low frequency signals to predict the ones of high frequency signals based on statistical correlation between high and low frequency signals, so as to reconstruct the lost high frequency signals[2-6]. And it needs no side information, is called "blind" bandwidth extension.

For audio signals, like music, full-band coding is usually employed, and the low and high frequency signals are encoded as a whole. This method has comparatively high bit rate, making it unsuitable for real-time network communication, only suitable for storage. In order to reduce the coding bit rate of audio signals, introducing bandwidth extension to reconstruct the high frequency parts in audio signals can attain low bite-rate encoding algorithms. However, considering the discrepancy between the statistical property of

ISSN: 2005-4254 IJSIP

audio signals and speech signals, "blind" speech bandwidth extension algorithms could not be applied to all kinds of audio signals. So it needs to add a little high frequency parameters side information on the basis of core encoders in audio bandwidth extension algorithm, using parameter encoding to rebuild the high frequency parts in audio signals [7-10]. This method is called “non-blind” bandwidth extension.

Liljeryd proposed the spectral bandwidth replication (SBR) method [11] for high frequency band reconstruction in 2002. SBR method uses a Quadrature Mirror Filter (QMF) bank to split the input signal into 64 sub bands, including 32 low frequency sub bands and 32 high frequency sub bands. The low frequency sub bands, which have high correlation with high frequency bands, are duplicated to high frequency, and are adjusted by using high frequency feature parameters to get the synthesized high-frequency signal. The method can reconstruct better high-frequency audio signals, and has been adopted by MPEG as a part of MPEG audio standards HE-AACv2. But the computational complexity of SBR is so high that the time consuming of SBR decoding is as high as the AAC core decoder [12]. The latest speech and audio mixed encoding standard of 3GPP is AMR-WB+ [13]. The BWE algorithm of AMR-WB+ extracts linear prediction coefficients (LPC) and sub-frame energy factor in time domain to reconstruct high frequency signal. The bit rate of the AMR-WB+ bandwidth extension algorithm is only 0.8kbps, which can satisfy the demands of audio codec in mobile communication. In 2009, AVS P10 proposed by AVS(audio and video standard in China) is a codec standard oriented to mobile communication. The bit rate of AVS-P10 bandwidth extension algorithm is also 0.8kbps, as same as AMR-WB+. And the AVS P10 has comparable coding quality compared with the BWE of AMR-WB+ [14].

For consumer electronic applications, especially for the devices working as networks terminals, such as audio surveillance, there are some constraints on implementation of BWE algorithm. These constraints includes: low computational complexity and low memory requirements; low bit-rate for network transmission; applicable to both music and speech preferably.

This paper proposed a mobile surveillance device oriented audio codec bandwidth extension algorithm. The rest of this paper is organized as followed: The 2nd part introduces our previous research work on bandwidth extension. The 3rd part introduces the proposed BWE coding and decoding algorithm. The 4th part conducts tests and analysis on new BWE coding algorithm. And the 5th part presents conclusion and future work.

2. Previous Work

Samsung and Wuhan University corporately proposed the bandwidth extension method for AVS P10 [14]. The basic principle of high frequency bandwidth extension in AVS-P10 standard is that utilizing the envelope information and gain parameters of high frequency signals to adjust the stimulus of low frequency signals and to reconstruct high frequency signals. The whole procession includes: extract parameters of frequency spectrum and sub-band gains of high frequency signals in encoder; utilize the parameters of high frequency envelope in decoder and the low frequency residual signal as stimulus signals to synthesize the high frequency basic signals, and then use the parameters of gains to adjust the high frequency basic signals, to generate the rebuilt high frequency signals in frequency domain, lastly transform the rebuilt high frequency signals in frequency domain to time domain to obtain reconstructed signals in time domain[15]. The encoding block diagram is shown in Figure1.

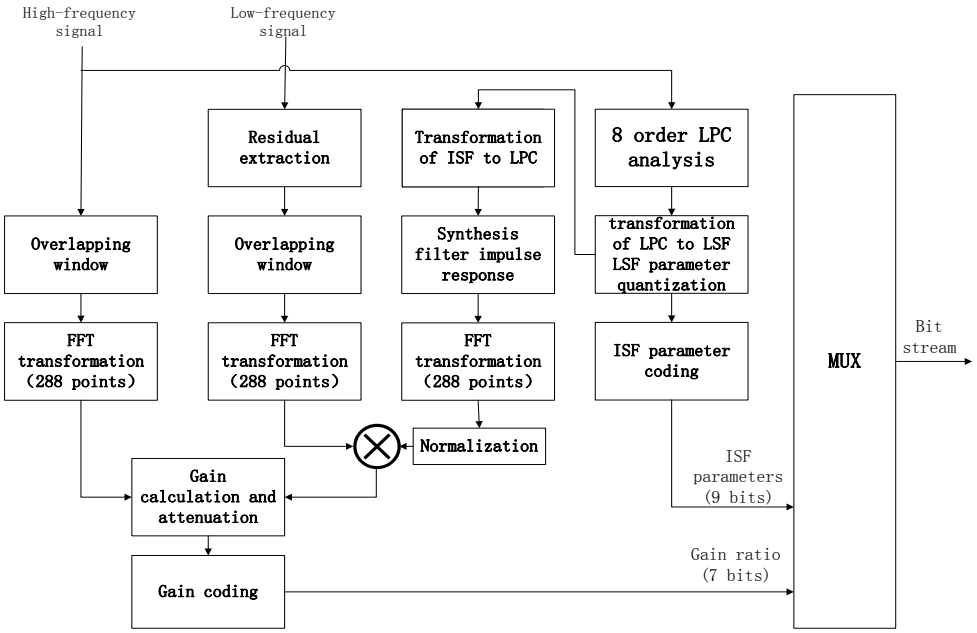

Figure 1. AVS P10 BWE Encoding Block Diagram

On the condition that the encoding bit rate of this method is equivalent to the one of AMR-WB+, the subjective quality of decoded audio sequences is comparable with the quality of AMR-WB+.

As introduced above, an advantage of "blind" BWE is the codec need no side-information, and "non-blind" BWE can get better audio quality with a higher bit-rate. In [16], we proposed an audio bandwidth extension algorithm, the main idea of the bandwidth extension method is combining the advantages of "blind" BWE and "non-blind" BWE: Firstly, a blind bandwidth extension method is used to remove the redundancy caused by the correlation between the low frequency and high frequency, and get the signal residual between the original signal and the signal predicted by blind bandwidth extension method. Then, a non-blind bandwidth extension method is used to encode the residual and transmit the encoding bit stream to decoder. The decoder decoding the residual signal and add the signal to the signal predicted by blind bandwidth extension method. At last, we can get the reconstructed high frequency signal. When predicting, we make a classification among the distinct audio signals to generate series of audio classified factors on which a previously trained mapping codebook was determined. Compared with AVS-P10, the proposed method can get comparable sound quality, but with lower bit-rate for high-frequency, and low computational complexity. This method can reduce the bit-rate, but can not improve the encoding quality for complex audio signals, such as the audio signal in surveillance.

3. Proposed Bandwidth Extension Codec

3.1. Framework of Encoder

Because of limitation in mobile surveillance device's computation and mobile channel bandwidth resources, the audio codec algorithm applied in mobile surveillance device should be characterized by low bit-rate and low computational complexity. Meanwhile, in consideration of the difference between the speech and music signals, as well as the different high frequency distribution attribute of different types of audio signals [17], to apply the same coding method for different types of signals will not achieve the perfect effect. In this paper, we propose an improved bandwidth extension method based on AVS P10 codec architecture. Because adopting the linear prediction could effectively remove

the speech signals' redundancy, therefore, for speech coding, we still adopt the existed AVS P10 bandwidth extension to reconstruct the high frequency signals. But linear prediction is not evident for musical signals' gains. Consequently, for music signals, we tend to employ new bandwidth extension method.

In AVS P10 low frequency core encoding algorithm, when it is speech-likely signal, we employ (ACELP) to encode; when it is music-likely signal, we employ (TVC) to encode. In the bit stream of core encoding, the adopted encoding modes will be marked, ACELP or TVC. So the encoding of high frequency could utilize this information: when low frequency encoding mode is ACELP, high frequency adopt the AVS-P10 BWE; when low frequency encoding mode is TVC, high frequency adopt the new bandwidth extension method. As illustrated in Figure2.

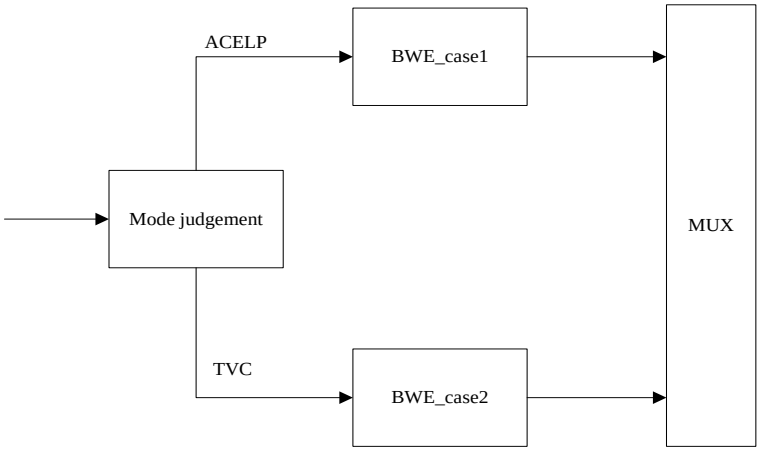

Figure 2. BWE General Framework

3.2. Coding Algorithm

1) BWE coding of speech like signal

When low frequency band adopt ACELP, high frequency coding will use AVS P10 BWE, we call it BWE_CASE1. The coding method is illustrated by Figure1.

2) BWE coding of music like signal

When low frequency band adopt TVC, it shows that the input is music like signal, BWE will use new method, and we call it BWE_CASE2.

The high-frequency parts of different types of audio signals have different features. For example, the high frequency part in percussion instrument is similar to noise, and the high frequency part in string instrument has obvious harmonic elements. Therefore, according to different types of signals, different high-frequency encoding modes are adopted, intra-frame low-frequency, inter-frame high-frequency signals or white noise are used to predict the high-frequency signal in current frame, and to get the most accurate encoding mode by comparing encoding SNR values, and extract high-frequency sub-band gain factors to reconstruct high-frequency signals in decoding part. This method can get the best predictive way which could effectively enhance the coding quality of audio high-frequency signal. In this case, we apply four predictive ways to reconstruct high-frequency signals, including translation and fold of low-frequency signals, high-frequency signal in previous frame, and white noise. In these four ways, utilizing fold and translation of low-frequency signals to rebuild high-frequency signals are two common bandwidth extension ways which can effectively reconstruct harmonic components. Sometimes, audio signals vary little in one period, and the high-frequency signals also change slightly. That is the high-frequency signal in current frame is similar to the one in previous frame, so using the frequency signal in previous frame to predict

the one in current frame can obtain less predicative error. Moreover, for some white-noise like high-frequency signal, white noise can be used to predict the current high-frequency signal. The proposed encoder is illustrated by Figure3.

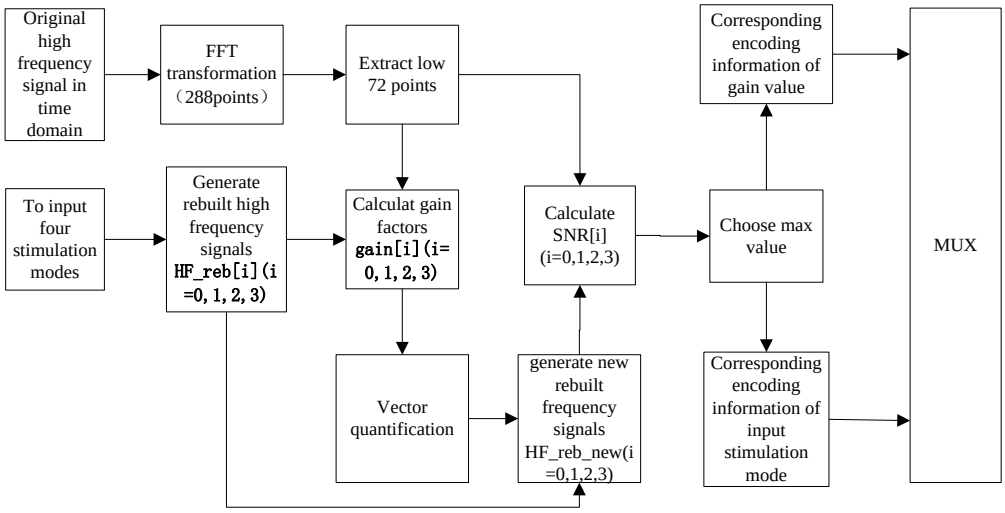

Figure 3. New BWE Encoder Block Diagram

When low frequency band adopt TVC, the process of high frequency bandwidth extension includes:

Make FFT on original high frequency, marked by HF_ori, divide the transformed frequency band into eight sub-bands. Then input four optional stimuli for reconstructing high frequency signals, include: low frequency residual signal, folded low frequency residual signal, high frequency signal of the previous frame, white noise; transform the four stimuli respectively to frequency domain, marked by HF_reb[i](i=0,1,2,3),and divide the transformed band into eight sub-bands; compute the proportional gain factors of the eight sub-bands, gain[i][j](i=0,1,2,3)(j=1,2,3,4,5,6,7,8), between the four stimuli and original high frequency signal HF_ori respectively, and quantify the generated gain factors,

Multiply the four stimuli respectively to corresponding quantified gain factors to obtain four rebuilt high frequency signal HF_reb_new[i](i=0,1,2,3). Then calculate the signal to noise ratios (SNR) of four high frequency generated by four different modes HF_reb_new[i](i=0,1,2,3) and original high frequency HF_ori, and then to choose the mode that has the maximum SNR value as the current stimulating one. The mode is coded by 2bits.

The selection of four stimulations is encoded by 2bits, respectively named by 11, 01, 10, 00. The stimulation modes and corresponding bits distribution are showed in Table1. Eight sub-bands' gain factors are divided into two four-dimensional vectors and each vector is quantified by 7bits.

Table 1. Four Stimulation Modes and Corresponding Coding

Rebuilt modes	Encoding of modes（2bits）
mode1	11
mode2	01
mode3	10
mode4	00

mode1: Fold the 4.8~6.4 kHz to 6.4~8 kHz to obtain rebuilt high frequency signal;
mode2: Copy 4.8~6.4 kHz to 6.4~8 kHz to obtain rebuilt high frequency signal;
mode3: Take white noise as rebuilt high frequency signal;
mode4: Take the high frequency signal in previous frame as the current rebuilt one.

3.3. Decoding Algorithm

1) BWE decoding of speech like signal

In decoding part, when the low frequency signals adopt ACELP to encode, high frequency would employ the method of AVS P10 to decode. Use low frequency residuals as stimulation, utilize the high frequency linear prediction coefficients to synthesis high frequencies, and employ the frequency spectrum energy gain factors to ultimately obtain rebuilt high frequency signal. As illustrated in Figure4.

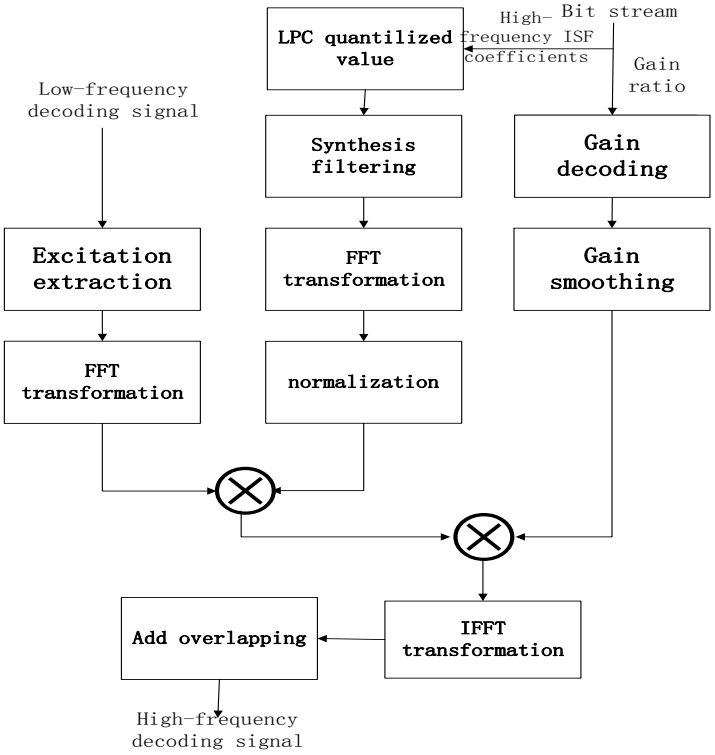

Figure 4. AVS P10 BWE Decoding Block Diagram

2) BWE decoding of music like signal

When the low frequency signals adopt TVC to encode, high frequency would employ new bandwidth extension to decode. Firstly, choose the best stimulation according to the decoded mode to obtain the reconstructed high frequency basic signals, and then to

reconstruct the high frequency signals based on the quantified gain factors of eight sub-bands as well as to smooth the frequency spectrum.

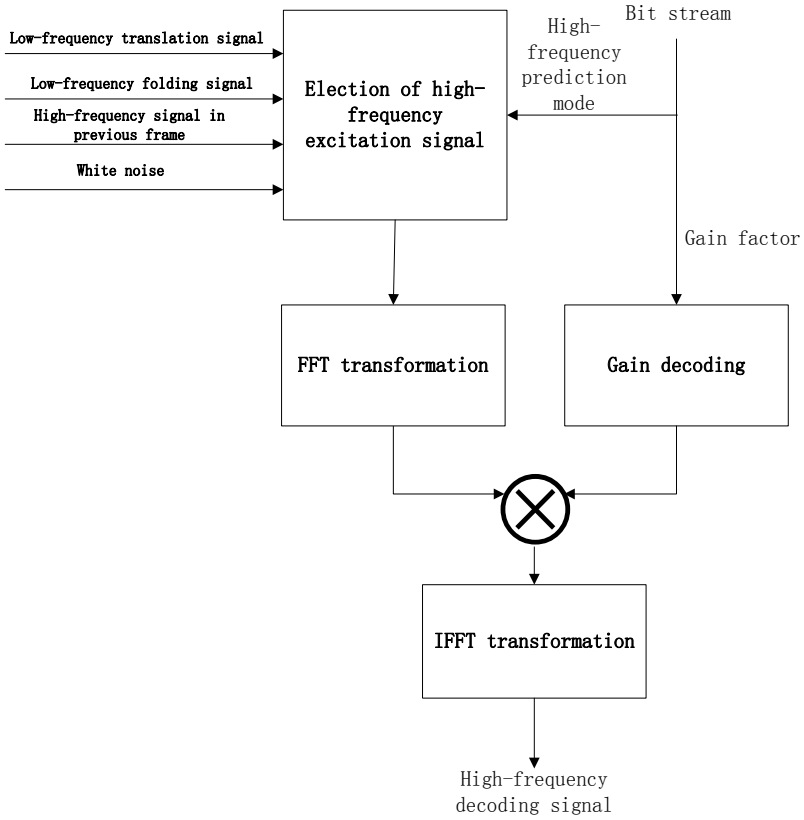

Figure 5. New BWE Decoding Block Diagram

As illustrated in Figure5, the decoding part firstly choose from copied and folded low frequency signals, high-frequency signal in previous frame and white noise to get high-frequency excitation of current frame, according to the high-frequency excitation reconstructed mode extracted from bit stream. Then quantified high-frequency gain factors are obtained from bit stream and codebook, and finally rebuilt high-frequency signals can be generated by both high-frequency excitation and gain factors.

3.4. Algorithm Implementation

In the process of algorithm implementation, we have taken some ways to make the computational complexity lower:

Firstly, White noises are average value calculated by large amount of audio high frequency signals instead of being generated randomly. Secondly, When the encoding and decoding of high-frequency signal in a frame finished, the high-frequency frequency-domain signal in Current frame are conserved in cache, and act as the previous frame high-frequency signal needed when the next frame signal is coded. Thirdly, Low frequency signals need to be transformed from time domain to frequency domain. And the signal needed by folded mode and copy mode both can be obtained from the low-frequency signal converted by time-frequency transformation. Therefore, no matter in encoder or decoder, obtaining all excitation signals needs time-frequency transformation for only once in each frame.

4. Experiment and Analysis

4.1. Subjective Test

In high quality audio subjective listening test, CMOS is commonly used. Each testing material consists of Ref/A/B, in which Ref is original uncoded signal, A/B are all decoded signals. In test, if A is the decoded signal results of proposed BWE method, then B is the decoded one for reference AVS P10; vice versa.

Specific testing method is: The position of Ref signal in every test is fixed, A and B is allocated randomly, and they are unknown to listeners. Listeners should be trained to be familiar with the whole testing process, and try to learn about some representative distortion presented frequently in testing materials. In high-quality marking system, 7 levels are used, shown as Table2, and the grades delivered by subjects should be integer.

Table 2. Levels Comparison Standard

Comparison of the Stimuli	Score
A is much better than B	+3
A is better than B	+2
A is slightly better than B	+1
A is the same as B	0
B is slightly better than A	-1
B is better than A	-2
B is much better than A	-3

The scoring rule is that the tested one which is much closed to the original one deserves high score, and the specific score is decided by the approaching degree.

(4)The testing result consists of average score and 95% confidence interval, and all testing results are needed to calculate statistical variance.

standard deviations： $\sigma = \sqrt{\frac{N\sum_{k=1}^{N} x_k^2 - (\sum_{k=1}^{N} x_k)^2}{N(N-1)}}$

95% confidence interval： $\mu \pm 1.96(\frac{\sigma}{\sqrt{N}})$

The testing sequences are 12 MPEG mono sequences designated by AVS with 16 kHz sample rate and 12 kbps coding bit-rate. This test includes 8 subjects who all possess acute hearing and have related working experience. The testing results are shown in Figure6.

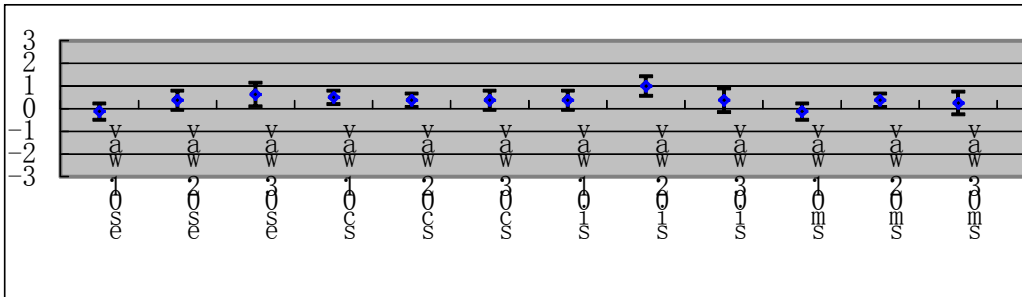

Figure 6. Subjective Testing Results. Positive Values Means the New BWE Method is better than AVS P10 BWE; Negative Values Means the New BWE Method is Worse than AVS P10 BWE

From the above subjective test, we can find that our new BWE method is better than BWE of AVS P10.

The Figure 7 shows the ratios of two cases in all test sequences.

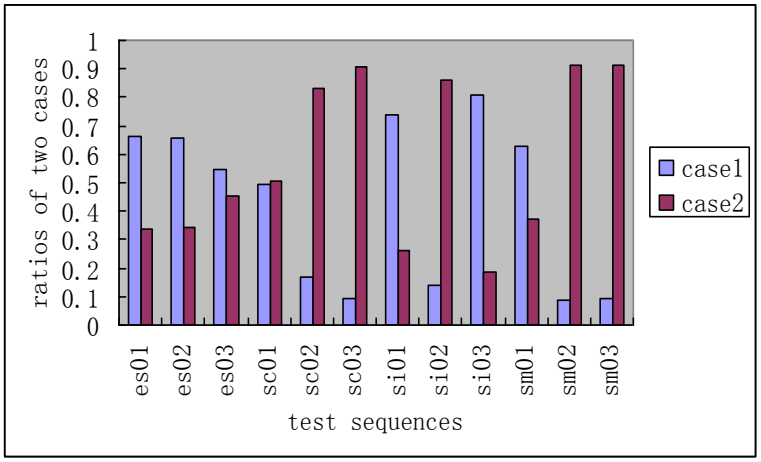

Figure 7. Comparison of Two Cases

From the Figure 6 and Figure 7, we can see the test sequences with high ratio of case 2, such as sc02, sc03, si02, sm02 and sm03, have better quality than original AVS-P10. That is because these sequences have more music-like frames and the case 2 method can encode signals better than case 1 method for music.

Complexity calculation: The testing sequences are continuously encoded and decoded for 30 times, both total codec time and time consumed on BWE related function are statistically calculated by Profiler tool in VC6. Test results are shown in Table 3:

Table 3. Complexity Calculation Statistics

	Original case	New case	Reducing rate
Encoding total time(s)	28890.012	27135.817	6.07%
Encoding BWE time(s)	5251.263	3707.537	29.40%
Encoding BWE rate	18.18%	13.66%	
Decoding total time(s)	6916.877	5924.494	14.35%
Decoding BWE time(s)	2401.653	1546.132	35.62%
Decoding BWE rate	34.72%	26.10%	

From the above data, we can find that the complexity on encoder of new BWE is reduced by 29.4% compared by the original one. For whole encoder, complexity is reduced by 6.07%. The main reason is that the FFT transformation in algorithm of case 2 in encoder is reduced by one time compared with the original one, and the calculation of high frequency signals linear prediction analysis and synthesis is also decreased by one time compared with original BWE method in AVS-P10. In decoder, there is 35.62% decrease in the complexity of BWE module , for the entire decoder, the complexity is reduced by 14.35%, the main reason is that in decoding part, the FFT time-frequency transformation in algorithm of case2 is lessened by one time, and there is no synthesis process in the new decoder.

4. Conclusions

BWE is one method that can enhance the quality of sound, which is especially attractive in consumer electronics. In this market, because of economic constraints on the size and cost of components, most manufacturers want to produce as cheaply as possible, yet retain a high subjective quality. This paper presents a BWE framework which provides different high-frequency BWE methods for speech like signal and music like signals. For different types of high-frequency part in audio signal, different high-frequency excitation reconstructed mode are applied, including intra-frame prediction by previous high frequency signal, inter-frame prediction by translational and folded low frequency signal, and white-noise prediction to choose the best prediction method to rebuild high-frequency signals. Subjective testing results demonstrate that the rebuilt quality of this proposed new method is better than the one in AVS P10, and obviously the computational complexity is reduced. This algorithm has been accepted by AVS and become a key technique for surveillance oriented audio codec standard AVS-S [18]. Our future work on this field is to find the reason of subject quality decline of a few sequences, such as es01 and sm01 and research better algorithm for case and mode selection.

Acknowledgements

The work in this paper cannot be achieved without the support of National Nature Science Foundation of China（61201247） and Nature Science Foundation of Hubei Province (2011CDB322).

References

[1] C. Avendano, H. Hermansky, and E. A. Wan, “Beyond NYQUIST: towards the recovery of broad-bandwidth speech from narrow-bandwidth speech”, In EUROSPEECH, **(1995)**, pp. 165-168.
[2] Y. M. Cheng, D. O'Shaughnessy, and P. Mermelstein, “Statistical recovery of wideband speech from narrowband speech”, Speech and Audio Processing, IEEE Transactions on, vol. 2, no. 4, **(1994)**, pp. 544-548.
[3] H. J. Liu, C. C. Bao and X. Liu, “Spectral envelope estimation used for audio bandwidth extension based on RBF neural network”, In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, **(2013)**, pp. 543-547.
[4] K. Li and C. H. Lee, “A deep neural network approach to speech bandwidth expansion”, In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on, **(2015)**, pp. 4395-4399.
[5] Y. Gu and Z. H. Ling, “Restoring high frequency spectral envelopes using neural networks for speech bandwidth extension”, In Neural Networks (IJCNN), 2015 IEEE International Joint Conference on, **(2015)**, pp. 1-8.
[6] Y. Wang, S. Zhao, Y. Yu and J. Kuang, “Speech Bandwidth Extension Based on GMM and Clustering Method”, In Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on, **(2015)**, pp. 437-441.
[7] E. Larsen and R. M. Aarts, “Audio Bandwidth Extension – application to Psychoacoustics, Signal Processing and Loudspeaker Design”, John Wiley & Sons, Ltd, Hoboken, New Jersey, USA, **(2004)**, pp. 154-161.

[8] X. Liu and C. C. Bao, "Audio bandwidth extension based on temporal smoothing cepstral coefficients", EURASIP Journal on Audio, Speech, and Music Processing, no. 1, **(2014)**, pp. 1-16.
[9] L. Jiang, R. Hu, X. Wang and M. Zhang, "Low Bitrates Audio Bandwidth Extension Using a Deep Auto-Encoder", In Advances in Multimedia Information Processing—PCM, Springer International Publishing, **(2015)**, pp. 528-537.
[10] T. Zernicki, M. Bartkowiak, L. Januszkiewicz and M. Chryszczanowicz, "Application of Sinusoidal Coding for Enhanced Bandwidth Extension in MPEG-H USAC", In Audio Engineering Society Convention 138. Audio Engineering Society, **(2015)**.
[11] M. Dietz, L. Liljeryd, K. Kjorling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding", In Audio Engineering Society Convention 112. Audio Engineering Society, **(2002)**.
[12] C. Dang, K. Dai, S. F. Wang, Y. Liu and Z. Y. Wang, "The Research and Implementation of SBR High Frequency Reconstruction Technology", Acta Electronica Sinica, vol. 32, no. 12, **(2004)**, pp. 189-191.
[13] "3GPP TS 26.290, Extended AMR Wideband codec", 3GPP, **(2004)**, pp. 39-41.
[14] "AVS-M2013, Low Complexity Bandwidth Extension", AVS 25th Meeting, Xiamen, China, **(2008)**.
[15] J. Zhan, K. Choo and E. Oh, "Bandwidth extension for China AVS-M standard", In Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, **(2009)**, pp. 4149-4152.
[16] B. Hang, R. Hu, Y. Yang and G. Gao, "A Low Bit Rate Mobile Audio High Frequency Reconstruction", In Audio Engineering Society Convention 129. Audio Engineering Society, **(2010)**.
[17] Y. T. Sha, C. C. Bao, M. S. Jia and X. Liu, "High frequency reconstruction of audio signal based on chaotic prediction theory", In Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, **(2010)**, pp. 381-384.
[18] "AVS-M2749, Bandwidth Extension Method for AVS-S", AVS 34th Meeting, Beijing, China, **(2010)**.

Authors

Bo Hang, He received the B.S. and M.S. degrees from Harbin Engineering University, Harbin, China, in 2000 and 2003, respectively, and the Ph.D. degree from Wuhan University, China, in 2012. He is currently an associate professor in Hubei University of Arts and Science, HBUAS. His research interests include multimedia compression, transmission, reproduction, and retrieval. He has led many national and international research projects and industrial projects.

Yi Wang, He received the B.S. degrees from Wuhan University, Wuhan, China, in 2002, and the M.E. degree from Wuhan University of Technology, China, in 2008. He is currently a lecturer in Hubei University of Arts and Science, HBUAS. His research interests include graph grammars, software engineering.

Changqing Kang, He received the B.S. degrees from Northeast Normal University, Changchun, China, in 2004 and the Ph.D. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, in 2009. He is currently an associate professor in Hubei University of Arts and Science, HBUAS. His research interests include signal enhancement and restoration, signal detection and extraction.