BANDWIDTH EXTENSION FOR CHINA AVS-M STANDARD

Jie Zhan, Kihyun Choo, and Eunmi Oh

Samsung Electronics Co., Ltd.
Mt. 14-1, Nongseo-dong, Giheung-gu, Yongin-si Gyeonggi-do, South Korea

ABSTRACT

We proposed a new frequency domain BandWidth Extension (BWE) technology. In the new technology, FFT based frequency domain gain shaping combined with Linear Prediction Coding (LPC) based spectral envelope shaping is used for generating high frequency signals. To preserve the amount of noise component in the reconstructed band, gain reduction controlled by Spectrum Flatness Measurement (SFM) is employed. Subjective testing results show that the presented technology exhibits a comparable performance compared to 3GPP AMR-WB+ with the same bit-rate in the framework of Audio Video coding of China Standard (AVS) Part 10 – Mobile Speech and Audio Codec. This technology has been formally adopted as the artificial high band coding module in AVS P10.

Index Terms— Audio Coding, LPC, Speech Coding, Standardization

1. INTRODUCTION

China AVS (Audio Video coding Standard) intends to standardize codecs with China's own intellectual property which is more economic for multimedia industries in China [1]. A Call for Proposal was issued on a universal speech and audio codec in 2005. The codec is officially named as AVS Part 10: Mobile Audio and Speech Codec [2]. The standardization aims at designing a mobile speech and audio codec to meet increasing demands in high quality mobile multimedia services such as China Mobile Multimedia Broadcasting (CMMB) system. The goal of AVS 10 is to achieve the reference quality of AMR-WB+ in an alternative way.

Fig. 1 shows the high level structure of AVS P10 decoder [3]. The input signal is split into Low Frequency (LF) band, and High Frequency (HF) band. The LF signal is coded with a switched Algebraic Code Excited Linear Prediction (ACELP) and Transform Vector Coding (TVC) module for coding mono signal, and a parametric stereo module for coding stereo signal. In HF band, the BWE technology described in this paper has been adopted which reduces bit rates by parametrically representing high band signal in encoder side and artificially reconstructing high band signal in decoder side.

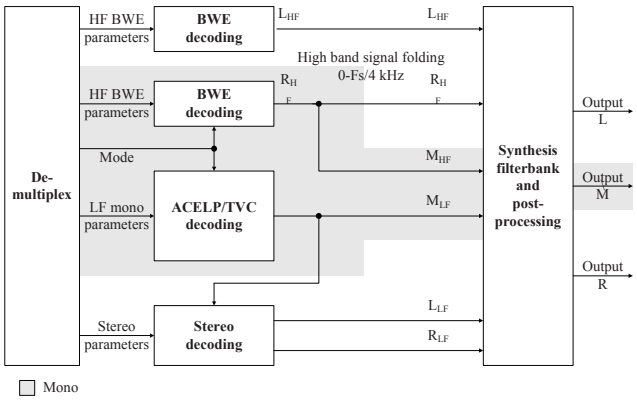

Fig. 1 High level decoder structure of AVS P10: Mobile Speech and Audio Codec

Bandwidth extension technologies are developed to enhance the coding efficiency in many audio and speech codec, for example, Spectral Band Replication (SBR) in mp3Pro/HE-AAC, spectral extension in Dolby Digital Plus, BWE in AMR-WB+ and so on. The idea of generating artificial high band signals from low band signal derives from perceptual and acoustic properties. Human hearing in high band is more sensitive to the spectral coarse structure than the spectral fine structure. Spectral fine structure in high band signal typically exhibits high correlation with that in low band [4].

The goal of this paper is to propose an alternative bandwidth extension technology without degrading the sound quality of AMR-WB+ under AVS P10 framework. The BWE in AMR-WB+ represents the high band signals in time domain by the spectral envelope and gain factors [5]. In comparison, the proposed technology operates in frequency domain to achieve good performance for music signals. In addition, gain reduction controlled by Spectrum Flatness Measurement (SFM) is introduced to preserve the amount of noise in high bands. This gain reduction makes the reconstructed sound more natural especially when the tonality of high bands differs from that of low bands.

The presented BWE technology uses a folding version of low band signal as a base signal. The spectrum of the

base band is adjusted to match the spectrum of an original high band signal by the high band LPC filter envelope information for each frame and the gain information for each sub-band. The SFM controlled gain reduction is employed to preserve the amount of noise component of the reconstructed high band signal in each sub-band, and gain smoothing is introduced to reduce gain discontinuity at the boundaries between neighboring sub-bands.

The subjective listening test results are given which exhibit that the performance of new technology is comparable to 3GPP AMR-WB+ BWE under AVS P10 framework. The proposed BWE has been accepted as the high band coding module in AVS P10 standard by China AVS [6][7].

2. FREQUENCY DOMAIN BWE ENCODER

The encoder structure of the presented frequency domain BWE is illustrated in Fig. 2. The inputs of BWE module are the high band signal and low band signal. The outputs of BWE module are the index for high frequency Immittance Spectral Frequency (ISF) and gain factors.

2.1. Base Signal Generation

One block of 2048-sample input waveform is low pass filtered and high pass filtered into low band and high band respectively. Down-sampling-by-two is performed to generate low band signal and high band signal consisting of 1024 samples each.

Eighth-order Linear Prediction Coding (LPC) coefficients of the high band signal are calculated by Levison-Durbin algorithm and transformed into ISF coefficients in each 256-sample frame. The HF ISF coefficients are quantized with 7 bits and transformed back into HF LPC coefficients. The 288-point impulse response of high frequency synthesis filter constituted by quantized LPC coefficients is transformed into frequency domain by FFT transformation to represent the spectral envelope of the high band signal.

Low band excitation signal is extracted by filtering the original low band signal through the low band LPC inverse filter. The low frequency LPC coefficients are innovated every 256-sample frame. Each 1024-sample super-frame of low band excitation signal is split into four frames by the cosine window, which has an overlap size of 32 samples between neighboring frames. Each 1024-sample super-frame of original high band signal is also windowed and split into four frames in the same way.

Each frame of the low band excitation signal and the impulse response of high band synthesis filter are transformed into the frequency domain by a 288-point FFT transformation. FFT coefficients are regrouped into real arrays in the order of: [X_0, $X_{N/2}$, $Re(X_1)$, $Im(X_1)$, ..., $Re(X_{N/2-1})$, $Im(X_{N/2-1})$], where N is the frame length, $Re(X_i)$ and $Im(X_i)$ represent the real and imaginary parts of i^{th} frequency bin respectively. X_0 and $X_{N/2}$ are real values.

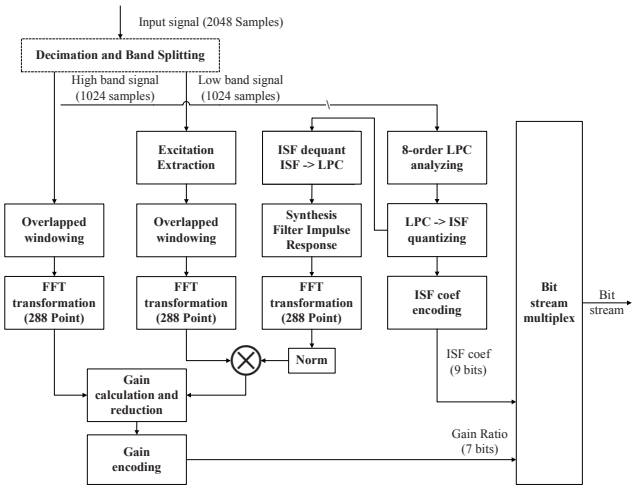

Fig. 2 Frequency Domain BWE Encoder Structure

The base signal is generated by multiplying the low band excitation spectrum with the normalized spectrum of high band synthesis filter impulse response. The operation is equivalent to modulating the spectral envelope of low band excitation by the high band spectral envelope.

2.2 Gain Calculation and SFM based Gain Reduction

Suppose that $X_{BF}[k]$ and $X_{HF}[k]$, $k=0...287$ represent FFT coefficients of the base signal and the high band signal in one frame, respectively. Each set of the FFT coefficients is further split into four sub-bands with uniform length of 72 coefficients. Gain factor is calculated in each sub-band by the equation (1)

$$Gain[i] = \frac{\sum_{j=0}^{M-1} (X_{HF}[iM + j])^2}{\sum_{j=0}^{M-1} (X_{BF}[iM + j])^2} \quad (1)$$

, where $Gain[i]$ denotes the gain factor of i^{th} sub-bands in current frame and M is the sub-band length, which is 72.

Although the high frequency spectral envelope can be coarsely represented by the HF LPC filter shaping and gain adjustment in most cases, the artificial high band signal may result in a murmuring sound when low band exhibits a noise-like feature while the high band exhibits a tonal-like feature. It usually happens in the case of the acoustic instrument, which has a strong high band harmonics, if the fine feature of the base band is not adjusted. It is mainly caused by tonality characteristic mismatch between the base signal and high band signal.

SFM is introduced to estimate tonal characteristics of base signal and high band signal [8] in each sub-band. It is

defined as the ratio between the geometric mean (GM) and the arithmetic mean (AM) of the power spectrum in logarithm domain, as described by equation (2).

$$SFM[i]_{\mathrm{dB}} = 10\log_{10}\frac{\sqrt[N]{\prod_{k=0}^{N-1}X^2(k)}}{\frac{1}{N}\sum_{k=0}^{N-1}X^2(k)} \quad (2)$$

SFM is non-positive. When it is zero, we can say that the spectrum shape is completely flat. When it reaches negative infinite, we can say the spectrum shape is not flat. By using SFM, tonal characteristic is adjusted to solve the problem of tonality mismatch between the original high band and the reconstructed high band.

Gain reduction is employed to reduce gain according to the ratio between the current base signal SFM factor and target high band SFM factor to preserve the amount of noise in reconstructed high band signal in the case that the noise-like low band spectrum is copied to the tonal-like high band without introducing additional bit consumption. The steps for gain reduction are:

- Calculate base signal SFM factor $SFM_{BF}[i]$ and high band SFM factor $SFM_{HF}[i]$ in each sub-band according to equation (1).
- If $SFM_{BF}[i] > SFM_{HF}[i]$, i.e, low band has a more noise like spectrum than target high band in the current sub-band, we calculate reduction factor $SFM_R[i]$ by equation (3). Otherwise, gain reduction is not applied. In this case, we set $SFM_R[i]$ to one.

$$SFM_R[i] = \frac{SFM_{HF}[i]}{SFM_{BF}[i]} \quad (3)$$

- Reduce the $Gain[i]$by multiplying with gain reduction factor $SFM_R[i]$in each sub-band according to equation (4).

$$Gain'[i] = SFM_R[i]\cdot Gain[i] \quad (4)$$

After the gain reduction, four reduced gain factors in each frame are vector quantized with 7 bits.

3. FREQUENCY DOMAIN BWE DECODER

The presented BWE decoder structure is illustrated in Fig. 3. The bit-stream is de-multiplexed, and low band excitation is decoded from core module. The high band ISF coefficients are decoded and transformed to LPC coefficients. The high frequency synthesis filter is derived from the decoded high band LPC coefficients. The impulse response of the high band synthesis filter is transformed into a frequency domain by FFT. The decoded low band excitation is transformed by FFT in the same way as the encoder side. The decoded low band excitation spectrum is multiplied by high band LPC synthesis filter impulse response spectrum to generate the reconstructed base signal. Gain indices are decoded and translated into gain factors. Decoded gain factors are smoothed and applied to the reconstructed base signal to generate the high band signal. Finally, overlapping-add is applied to generate the decoded high band waveform.

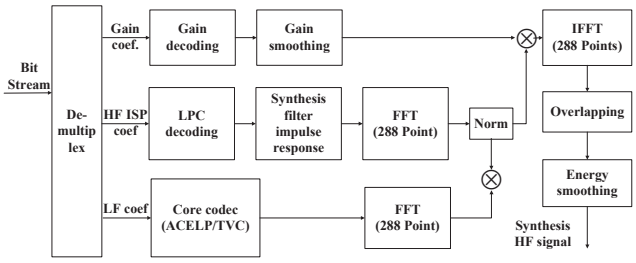

Fig. 3 Frequency Domain BWE Decoder Structure

In each frame, gain indices are decoded into 4 gain factors $\hat{G}ain[i]$. Before applying gains to the reconstructed base signal, they are smoothed by linear interpolation. The gain smoothing is introduced to reduce discontinuity at the boundary between neighboring sub-bands [6]. The interpolated gain is calculated by the equation (5), where, $Gain^i(k)$ represents the smoothed gain in k^{th} position of i^{th} sub-band, N is sub-band length. We assume that $\hat{G}ain[0] = \hat{G}ain[1]$ and $\hat{G}ain[4] = \hat{G}ain[3]$.

$$Gain^i(k) = \begin{cases} \hat{G}ain[i]*\frac{\frac{N}{2}+k}{N} + \hat{G}ain[i-1]*\left(1-\frac{\frac{N}{2}+k}{N}\right), 0 \le k < \frac{N}{2} \\ \hat{G}ain[i+1]*\frac{\frac{N}{2}-k}{N} + \hat{G}ain[i]*\left(1-\frac{\frac{N}{2}-k}{N}\right), \frac{N}{2} \le k < N-1 \end{cases} \quad (5)$$

Smoothed gains are applied to the base signal $\hat{X}_{BF^i}(k)$to generate the artificial high band spectrum $\hat{X}_{HF^i}(k)$ by

$$\hat{X}_{HF}{}^i(k) = \hat{X}_{BF}{}^i(k) * Gain^i(k) \quad (6)$$

Inverse FFT transform are applied to reconstruct high band signals in each frame. Four frames are overlapped and added to recreate one super-frame.

4. PERFORMANCE ANALYSIS

The performance of BWE technology is measured mainly by subjective quality measurement. In our testing, Comparison Mean Opinion Score (CMOS) is used, the test material consists of Ref/A/B, in which Ref is the original un-coded sample, both A and B are decoded signals. In each test, when A is decoded result of AVS P10 with AMR-WB+ BWE, B is the decoded result of AVS P10 with our new BWE, or vice versa. The score has 7 levels, which are listed in Tab.1. Six listeners performed the listening test. All of them are experts in the audio field.

Tab.1. Levels comparison standard.

Comparison of the Stimuli	Score

B is much better than A	+3
B is better than A	+2
B is slightly better than A	+1
B is the same as A	0
B is slightly worse than A	-1
B is worse than A	-2
B is much worse than A	-3

The test results consist of an average score and a 95% confidence interval. We use 23 AVS testing items as the testing material. All of them are a 48 kHz sampled mono waveform. Testing bit-rates are 12kbps and 24kbps.

The average listening test results for 12kbps and 24kbps are shown in Fig. 4. The test results for each item at 12kbps and 24kbps are given in Fig. 5 and Fig. 6 respectively. From Fig. 4, on average, the qualities of our BWE at both 12kbps and 24kbps are statistically comparable to that of AMR-WB+ BWE in 95% confidence interval sense.

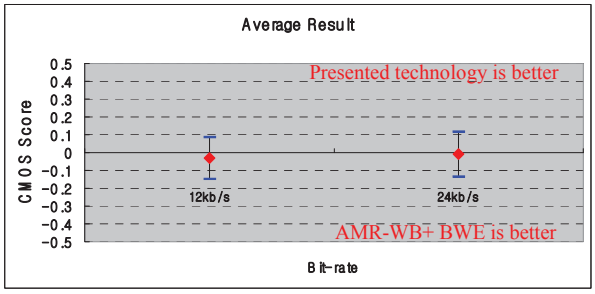

Fig. 4 Average CMOS score at 12kbps and 24kbps Mono

At 12kbps mono, all the cases are statistically comparable to that of AMR-WB+ BWE in a 95% confidence interval sense. In the 5 cases, the average scores are higher, and in 10 cases, the average scores are lower than that of AMR-WB+. The others are the same.

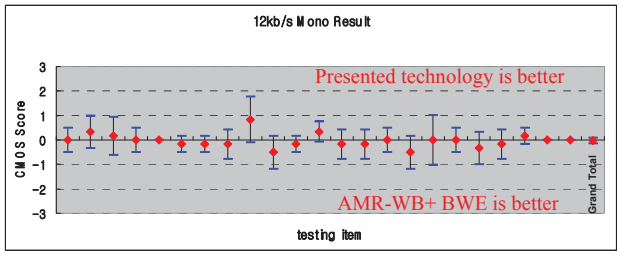

Fig. 5 CMOS score of each item at 12 kbps mono

At 24kbps mono, all the cases are statistically comparable to that of AMR-WB+ BWE in 95% confidence interval sense. In 6 cases, the average scores are higher, and in 8 cases, the average scores are lower than that of AMR-WB+. The others are the same.

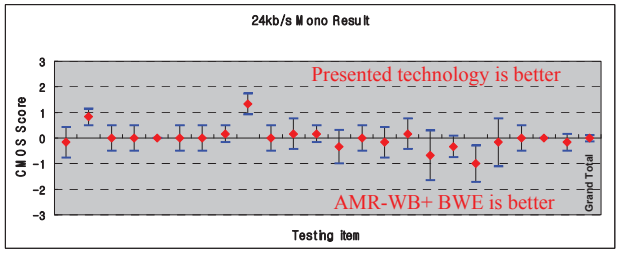

Fig. 6 CMOS score of each item at 24 kbps mono

5. CONCLUSION

Subjective listening test results demonstrate that the performance of the presented technology is comparable to 3GPP AMR-WB+ BWE under AVS P10 framework in an alternative approach. The proposed technology utilizes SFM controlled gain reduction and gain smoothing. This technology has been formally adopted in high band coding of AVS P10 for mobile applications.

6. REFERENCES

[1] L. Miao, S Kim, S Lee and J Van, "Context-dependent bitplane coding in China AVS audio," *Proc. IEEE ISPACS-2005*, pp. 765-768, 2005, December

[2] AVS-N1233, "AVS Mobility Audio Codec Call For Proposal," *AVS 15th Meeting*, Qingdao, China, 2005.12.

[3] AVS-M2421, "AVS P10: Mobile Speech and Audio Codec Committee Draft v1.0," *AVS Beijing Ad-hoc Meeting*, Beijing, China, 2008.07.

[4] M. Dietz, L. Liljeryd, K. Kjorling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," *112th AES Convention, Munich, Germany*, 2002, May

[5] J. Makinen, B. Bessette, S. Bruhn, P. Ojala, R. Salami, and A. Taleb, "AMR-WB+: A New Audio Coding Standard For 3rd Generation Mobile Audio Services," *Proc. IEEE ICASSP-2005*, pp. 1109-1112, vol. 2, 2005. March

[6] AVS-M2013, "Low Complexity Bandwidth Extension," *AVS 25th Meeting*, Xiamen, China, 2008.6

[7] Y. Zhang, R. Hu, G. Gao, "Artificial Mobile Audio Bandwidth Extension," *Communications and Information Technologies*, 2006. ISCIT '06. International Symposium on, pp. 410-413, 2007.09

[8] J. Johnston, "Transform coding of audio signals using perceptual noise criteria," *Selected Areas in Communications, IEEE Journal on, 6 (1988)*, pp.314–323.