SUPERWIDEBAND EXTENSION FOR AMR-WB USING CONDITIONAL CODEBOOKS

Yingxue Wang, Shenghui Zhao, Kadri Mohammed, S.Danish Abbas Bukhari, Jingming Kuang

School of Information and Electronics, Beijing Institute of Technology,
Beijing 100081, China

ABSTRACT

In this paper, the conditional codebook mapping (CCM) method is proposed for super-wideband extension (SWB) of speech signals to achieve a wider bandwidth up to 12.8 kHz. As we know, the performance of bandwidth extension by codebook mapping (CM) is restricted by the codebook size. The proposed CCM-based BWE method reduces the restriction while introducing a conditional codebook and is applied in AMR-WB at 19.85 kbps. Both the Objective test results in the term of Log-spectral distortion (LSD) and Perceptual Evaluation of Speech Quality (PESQ) and the subjective test results show that the CCM-based SWB extension method achieves higher speech quality compared with the CM method.

Index Terms—Bandwidth extension, super-wideband extension, codebook mapping, conditional codebook mapping

1. INTRODUCTION

As we know, adaptive multi-rate wideband (AMR-WB) is designed to provide a speech bandwidth of 7 kHz to obtain higher speech quality over the narrowband speech codecs, e.g., adaptive multi-rate narrowband (AMR-NB). And the speech bandwidth is very important for the upper bound of the reconstructed speech quality. Hence it is necessary to extend the wideband to a wider bandwidth (i.e., super-wideband). In fact, the super-wideband (SWB) extension is experiencing a strong momentum in today's speech communication systems.

There has been a significant amount of bandwidth extension (BWE) technologies to achieve SWB extension. Four classifications of BWE methods can be made according to the different criteria, including blind BWE and the non-blind BWE method [1], source-filter model (SFM) and non-source-filter model (NSFM) [2], time-domain BWE algorithm and transform-domain [3], BWE algorithm based on decoding and BWE algorithm based on information hiding [4]. Most of the bandwidth extension (BWE) algorithms proposed in the literature are based on the source-filer mode (SFM) of high frequency production which can be divided into two parts: the excitation extension and the spectral envelop extension. Among these SFM-based BWE methods, codebook mapping (CM) method is commonly used. However, one restriction of CM method is that the codebook size which can be increased to improve the performance while resulting in the increment of the bit rate. In order to reduce this restriction, we propose the CCM-based BWE method in which the bit rate is not up to the original codebook size but the conditional codebook size.

This paper is organized as follows. Section 2 gives an overview of AMR-WB codec while section 3 presents the CM method in AMR-WB at 19.85 kbps. In section 4, the details of the encoder and the decoder are descripted respectively. Section 5 presents experimental results and performance comparison. Section 6 concludes this paper.

2. AN OVERVIEW OF AMR-WB

The AMR-WB codec has been standardized by 3GPP and ITU-T for wideband speech conversational applications in March 2001 and in July 2002 respectively. It is a multi-rate speech codec based on Algebraic Code Excited Linear Prediction (ACELP) with 9 different bit rates which are 6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 and 23.85 kbps [5]. It processes 16 kHz sampled input signals with the internal sampling rate of 12.8 kHz, and provides a bandwidth of 0.05-7 kHz. The frame size of AMR-WB is 20 ms.

A BWE module based on the SFM method is introduced in 23.85 kbps mode of AMR-WB to produce a bandwidth of 6.4-7 kHz. Although the BWE method in AMR-WB achieves good speech quality, it is not suitable for the SWB extension. We hence propose a CCM-based SWB extension method to extend the bandwidth of the speech up to 12.8 kHz.

3. THE CONVENTIONAL CM-BASED SWB EXTENSION

In this section, we give an overview of the conventional CM-based BWE method and outline the corresponding SWB extension scheme in AMR-WB.

The conventional CM-based BWE generally consists of two modules, namely high-band excitation signal estimation module and high-band spectrum envelope estimation module. High-band spectral envelop extension can be formulated as $H = C\{L\}$, where L is the feature vector of low-band spectral envelope while H is the feature vector of high-band spectral envelope, $C\{\bullet\}$ is the Codebook Mapping (CM) or Weighted Codebook Mapping (WCM). It is can be seen from the above formulation that the value of H is decided by both the feature vector of low-band spectral envelope L and $C\{\bullet\}$, so selecting the appropriate low-band characteristic parameters and the proper codebook is vital. Currently, the most commonly used L is in form of Liner Prediction Cepstral Coefficients (LPCCs), Cepstral Coefficients (CEPs), Mel Frequency Cepstral Coefficients (MFCCs) [6] [7], the Line Spectral Frequencies (LSFs) [8] and Immittance Spectral Pairs (ISP) [9].

Fig. 1 shows the basic structure of the CM-based SWB extension in AMR-WB. It can be seen that the encoder and decoder consist of two parts, i.e., the AMR-WB core codec and the enhancement layer. The AMR-WB codec is used to encode and decode the low frequency while the enhancement layer is based on the conventional CM-based SWB extension scheme. The enhancement layer codec has the same frame structure with that of the AMR-WB codec, operating on per frame of 20 ms with a sampling rate of 32 kHz. In Fig. 1, we can see that two important parameters are the gain index and the LF ISF index, which are both transmitted from the encoder to the decoder. The gain index and the LF ISF index are responsible for obtaining the gain factor which is used to scale the low frequency excitation in order to regenerate the HF excitation and reconstructing the HF spectral envelop respectively. The gain index is obtained by the conventional vector quantization. The LF ISF index is obtained by comparing the input LF ISF vector with the code-vectors using the mean square measure in the one-to-one codebook, so called LF_HF_ISF codebook which is designed using the conventional LBG algorithm. With the HF excitation and the HF spectral envelop, the HF signal is reconstructed by filtering this HF excitation through the HF synthesis filter. The enhanced HF synthesized signal together with the AMR-WB synthesis is up-sampled and filtered through the low-pass filter to form the 32 kHz sampled output signal, offering a bandwidth ranging from 50 Hz up to 12.8 kHz.

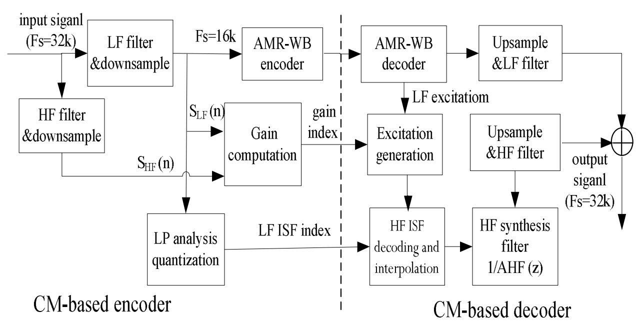

Fig. 1. The structure of the CM-based SWB extension encoder and decoder

4. THE CCM-BASED SWB EXTENSION

4.1. Principle of the CCM-based SWB extension

The proposed CCM-based BWE method is based on the conventional CM-based method, and the main difference is that the CCM-based BWE method introduces the conditional codebooks which are composed of an one-to-one codebook from LF spectral envelop parameters to HF spectral envelop parameters, so called LF_HF_ISF and a set of HF spectral envelop parameters codebooks named HF_ISF codebook.

The three parameters needed to be extracted are gain index, HF IF index and LF ISF index. Among the three parameters, only gain index and HF IF index are transmitted in the CCM-based SWB extension scheme under the assumption that the LF ISF index at the encoder is identical to the LF ISF index at the decoder. The gain index is still used to reconstruct the HF excitation, whereas the HF IF index instead of LF ISF index is used to regenerate the HF spectral envelop by finding the appropriate HF ISF vectors in conditional codebook while the LF ISF index acts as a medium, aiming at finding out the corresponding HF ISF codebook. The LF_HF_ISF codebook is firstly searched using the mean square measure, returning the optimum LF ISF index. The HF_ISF codebook is then searched for a better match followed by the corresponding HF ISF index being transmitted to the decoder.

The LF_HF_ISF codebook is first designed using the conventional LBG algorithm, from which the HF_ISF codebooks are subsequently generated by only using a subset of the code vectors in the LF_HF_ISF codebook.

We give the encoder and decoder structures of the proposed CCM-based SWB extension in AMR-WB codec as below.

4.2. The encoder structure

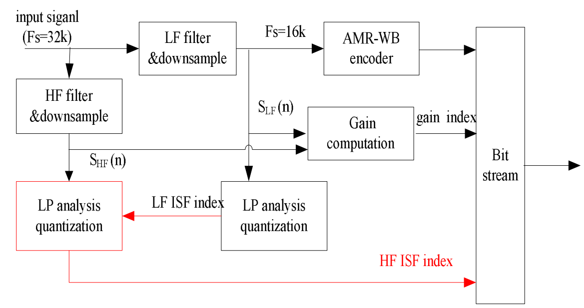

Fig. 2. The structure of CCM-based SWB extension encoder

Fig. 2 shows the basic structure of the CCM-based SWB extension encoder. Firstly, the input signal sampled at 32 kHz is filtered through a low-pass filter and a high-pass filter followed by passing through a re-sampler to resample the low frequency signal(i.e., 0.05-7kHz) and the SWB signal (i.e., 7-16 kHz) to two 16 kHz sampled signals $s_{LF}(i)$ and $s_{HF}(i)$. Then, for every frame, the low frequency signal $s_{LF}(i)$ and high frequency signal $s_{HF}(i)$ are individually analyzed to extract the corresponding 16*th*-order spectral coefficients LF ISF and the 8*th*-order HF ISF. Then, we firstly extract the LF ISFs followed by the quantization of LF ISFs, i.e., comparing the current frame LF ISF vector with the first 16-dimension LF ISF vectors in the one-to-one LF_HF_ISF codebook, and obtain the LF ISF index $m \in \{1,2,\cdots M\}$ which will be used to choose the corresponding HF_ISF codebook among M HF_ISF codebooks. Then the HF ISF vector is extracted followed by the quantization of HF ISF vector through comparing the current frame HF ISF vector with the HF ISF vectors in the corresponding HF_ISF codebook, giving the HF ISF index of the best matching vector which will be transmitted as the side information to the decoder. Meanwhile, the ratio between the energy of $s_{LF}(i)$ and the energy of $s_{HF}(i)$ which is denoted as gain G_i is computed every sub-frame. Consequently, the quantization of $\log_{10} G_i$ vector for each frame is carried out, giving the index of $\log_{10} G_i$ vector which in combination with the HF ISF index is transmitted as the side information to extend the bandwidth up to 12.8 kHz.

As described above, when encoding the input vector, it should be noted that it is required to store a gain codebook and a conditional codebook before the quantization. The gain codebook is a conventional codebook designed using the LBG algorithm. The structure of the conditional codebook is illustrated in Fig. 3. The 24-dimensional one-to-one LF_HF_ISF codebook consists of M vectors. For each sub-vector, the first 16-diemsional coefficients are denoted as the LF ISF while the last 8-diemsional coefficients are denoted as the corresponding HF ISF. The 8-dimensional HF_ISF codebook consists of N vectors.

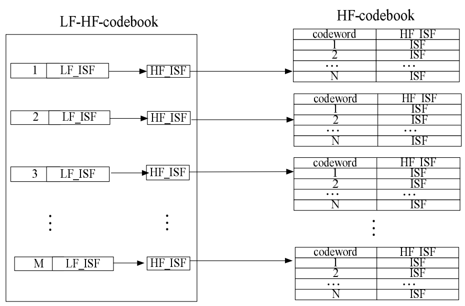

Fig. 3. The structure of the conditional codebook

4.3. The decoder structure

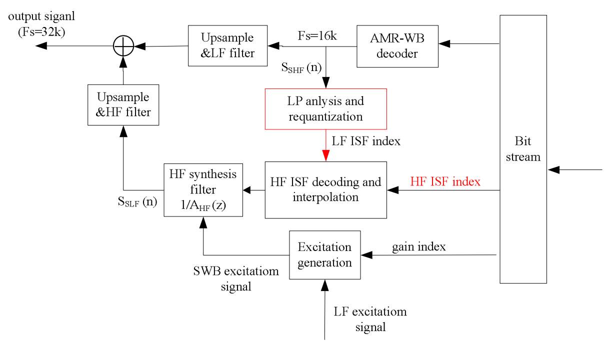

Fig. 4. The structure of the CCP-based SWB extension decoder

Fig. 4 gives the basic structure of the CCM-based SWB extension decoder. At the decoder, the gain indices and the HF ISF indices are received from AMR-WB bit-stream while the LF ISF index is obtained by comparing the ISF vector of the decoded LF signal with the LF ISF vectors in the LF_HF_ISF codebook. The transmitted HF ISF index is used to retrieve the HF spectral parameters while the LF ISF index is used for selecting the corresponding HF_ISF codebook among a set of HF_ISF codebooks. With the gain index, the gain factor k used to scale the low band excitation $e(i)$ is given by

$$k = \sqrt{\frac{G_{o1}}{G_i}},$$

where G_{o1} is the gain between the energy of the synthesis low frequency and that of the output signal obtained with the low frequency excitation $e(i)$ passing through the high frequency liner synthesis filter. The gain-scaled SWB excitation signal $e_h(i)$ is given by

$$e_h(i) = ke(i).$$

5. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed CCM-based SWB extension scheme, we carried out objective test, subjective test and complexity test, while the bit rate is also examined. In all cases, we use 50 training speech samples consisting of 25 male sentences and 25 female sentences, and 20 test speech samples consisting of 10 male sentences and 10 female sentences which are all 32 kHz sampled. What is more, the LF_HF_ISF codebook size for CMM-based SWB method is 4 for all tests below. Note that, theoretically, LF_HF_ISF codebook can be of any size. However, increasing the LF_HF_ISF codebook size would result in more HF_ISF codebooks which would lead to the increased computational complexity and memory requirement.

5.1. Objective evaluation

Log Spectral Distance (LSD) and Perceptual Evaluation of Speech Quality (PESQ) are adopted in the objective tests respectively. Twelve male speech samples and twelve female speech samples are involved. In addition, the LSD measure is applied to the high-frequency band (7-12.8 kHz) between wideband and the estimated speech. The lower the LSD value means the higher speech quality. In the LSD test, the numbers of bits transmitted $(n = 2,3,4,5,6,7,8,9)$ are exploited as the conditions.

As to the PESQ test, the output of the proposed scheme is firstly down-sampled to 16 kHz. The HF_ISF codebook size is 512. PESQ scores are ranging from 1.0 (lower relevance) to 4.5 (greater relevance). Table 1 shows the PESQ scores while Fig. 5 gives the results by LSD measure.

Table 1. PESQ test result (Mean scores)

Speech File	Codec	
	AMR-WB at 19.85 kbps	CCM-based method in AMR-WB at 19.85 kbps
Male	3.904	3.972
Female	3.869	3.977
Average	3.886	3.975

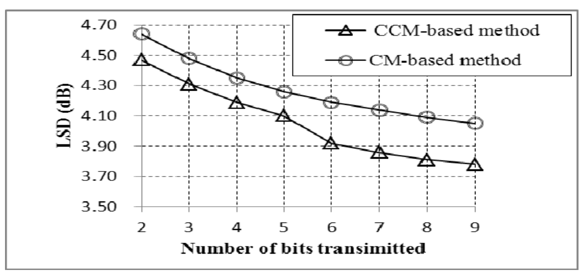

Fig. 5. LSD results for high frequency band

As can be seen from Table 1, the proposed CCM-based SWB extension scheme achieves higher PESQ scores, which means that the proposed CCM-based SWB extension scheme in AMR-WB at 19.85 kbps obtains higher speech quality than that of the original AMR-WB at 19.85 kbps. Figure 5 shows that the CCM-based SWB method achieves lower LSD values (approximately 0.14 dB mean improvement) compared with the traditional codebook mapping method with the same number of bits transmitted. Meanwhile, the proposed CCM-based SWB method achieves the same LSD value as that of the conventional CM-based method with a lower bit rate. In other words, the CCM-based BWE method can obtain the same speech quality with less number of bits need to be transmitted.

5.2. Subjective evaluation

To evaluate the subjective quality of the proposed CCM-based method, A-B preference tests (A-speech by CCM-based SWB method, B-speech by CM-based SWB method) are carried out and 12 subjects are involved in the test. Fig. 6 shows results of the A-B preference tests.

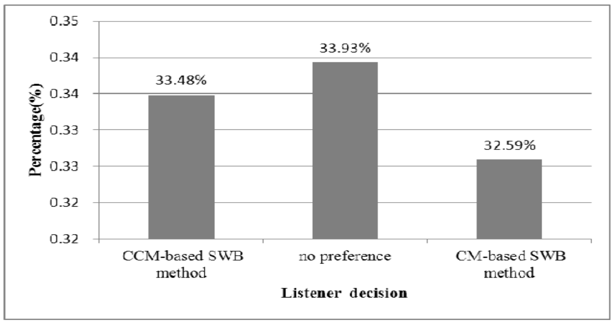

Fig. 6. Subjective comparison test result

As can be seen from Fig. 6, CCM-based SWB extension in AMR-WB at 19.85 kbps achieves higher performance than the conventional CM-based SWB extension.

5.3. Complexity test and bit rate

Usually, the computational complexity is specified in units of Weighted Million Operations Per Second (WMOPS) required by per frame. As shown in Table 2, the total complexity of the enhancement layer is 3.88 WMOPS and the overall complexity of the proposed scheme is 42.78 WMOPS. The computational complexity caused by the CCM-based SWB extension is low and since we spend most of time searching the HF_ISF codebook, the computational complexity is almost same as that of the conventional CM-based method.

Table 2. Complexity comparison test result

	En-/decoder	Complexity (WMOPS)
AMR-WB codec	encoder	31.1
	decoder	7.8
The enhancement layer	encoder	2.71
	decoder	1.17

The CCM-based SWB extension scheme has a total number of $\log_2 N + \log_2 W$ bits. Overall, the increased bitrate due to CCM-based SWB extension is $\left. \log_2 N + \log_2 W \middle/ 20 \right.$ kbps.

6. CONCLUSION

A SWB extension method based on conditional codebook mapping (CCM) is proposed for AMR_WB. Both the objective tests and the subjective tests show that it achieves improvement of speech quality over the conventional CM-based method. In the next step, we will integrate data hiding with SWB extension, aiming at maintaining the same bit rate while improving the speech quality.

7. ACKNOWLEDGE

The work in this paper is supported by the research cooperation between Beijing Institute of Technology (BIT) and Ericsson.

8. REFERENCES

[1] Nakatoh, Yoshihisa, M. Tsushima, and Takeshi Norimatsu. "Generation of broadband speech from narrowband speech using piecewise linear mapping", in *Proc. EuroSpeech*, vol. 3, pp. 1643-1646, 1997.

[2] Jax, Peter, and Peter Vary. "On artificial bandwidth extension of telephone speech", *Signal Processing*, pp. 1707-1719, 2003.

[3] Tsujino, Kosuke, and Kei Kikuiri. "Low-complexity bandwidth extension in MDCT domain for low-bitrate speech coding", in *Proc. ICASSP*, pp. 4145-4148, 2009.

[4] Geiser, Bernd, Peter Jax, and Peter Vary. "Artificial bandwidth extension of speech supported by watermark-transmitted side information", in *Proc. InterSpeech*, pp. 1497-1500, 2005.

[5] Bessette, Bruno, et al. "The adaptive multirate wideband speech codec (AMR-WB)", *IEEE Transactions on Speech and Audio Processing*, pp. 620-636, 2002.

[6] A.H. Nour-Eldin and P. Kabal, "Mel-frequency cepstral coefficient-based bandwidth extension of narrowband speech", in *Proc. InterSpeech*, pp. 55-56, 2008.

[7] Seltzer, Michael L., Alex Acero, and Jasha Droppo. "Robust bandwidth extension of noise-corrupted narrowband speech", in *Proc. InterSpeech*, pp. 1509-1512, 2008.

[8] Chennoukh, S., et al. "Speech enhancement via frequency bandwidth extension using line spectral frequencies", in *Proc. ICASSP*, pp. 665-668, 2001.

[9] Bistritz, Yuval, and Shlomo Peller. "Immittance Spectral Pairs (ISP) for speech encoding", in *Proc. ICASSP*, pp. 9-12, 1993.