Proceedings of the 2012 2nd International Conference on Computer and Information Application (ICCIA 2012)

An improved method for AMR-WB speech codec

Chunling Zhang, Shenghui Zhao, Hongyuan Xiao, Jing Wang, Jingming Kuang
School of Information and Electronics, Beijing Institute of Technology, Beijing, 100081, China
e-mail:10905012@bit.edu.cn

***Abstract*—an improved method is proposed to skip the look-ahead period in this paper. The improved method uses the autocorrelation algorithm to calculate the Linear Prediction (LP) coefficients and then the LP coefficients are employed to extrapolate new samples for replacing the look-ahead samples. To evaluate the quality of this method, perceptual evaluation of speech quality (PESQ) and the A/B listening test method are designed for the objective evaluation and subjective evaluation. The reconstructed quality of the modified method is near to the original AMR codec, at the same time, the delay of the improved method is lower 5ms than the original method.**

Keywords- AMR, speech codec, autocorrelation, look-ahead，Linear Prediction

I. INTRODUCTION

In 1999 3GPP together with ETSI (European Telecommunication Standards Institute) started development and standardization of a wideband speech codec for the WCDMA 3G and GSM systems. After almost two years of intense development and two competitive codec selection phases, the wideband codec algorithm was selected in December 2000. The speech codec specifications were finalized and approved in March 2001. The codec is referred to as Adaptive Multi Rate Wideband (AMR-WB) codec. This codec has been approved by the ITU in January 2002 and it is known as ITU-T Recommendation G.722.2 [1] [2].

The AMR-WB codec is the voice codec standardized for GSM and WCDMA 3G systems, which has been developed for use in several applications: the GSM full-rate channel, the GSM EDGE Radio Access Network (GERAN) 8-Phase Shift Keying (8-PSK) Circuit Switched channels, the 3G Universal Terrestrial Radio Access Network (UTRAN) channel and also packet based voice over internet protocol (VoIP) applications. The adoption of AMR-WB is of significant importance since for the first time the same codec is adopted for wireless as well as wire line services[3] [4].

The AMR-WB speech codec is based on Algebraic Code Excited Linear Prediction (ACELP) technology. In the coding process, speech samples of the next frame are utilized during the Linear Prediction LP analysis of the current frame, [5]. The time spent on waiting next frame samples is referred to the look-ahead period, which will cause an extra delay. Delay is an important concern for real-time two-way conversations, and it basically can be thought of as the time the sound takes to travel from speaker to listener. For an excessively large delay, that is, above 150ms, the ability to hold a conversation is impaired. The parties involved begin to interrupt or “talk over” each other because of the time it takes to realize the other party is speaking. When delay becomes high enough, conversation degrade to a half-duplex mode, taking place strictly in one direction at a time; hence, the lower the delay the better.

For AMR-WB, the frame is 20ms and every frame has 4 sub-frames, therefore each subframe is 5ms and the look-ahead period is 5ms. Just like AMR-WB, AMR-NB[6],AMR-WB+[7], AVS[8], SILK[9] and some other codecs have the same problem. However, few research works about skipping the look-ahead samples have been published. In this paper, we propose a method to skip the look-ahead period without significantly affecting the quality of the coded speech. The remainder of the paper is organized as follows: In section 2, the extrapolated method is described in detail. Then Section 3 gives the subjective experimental results and the conclusion is presented in section 4.

II. The ORIGINAL method

The sampling rate of AMR-WB codec is 16 kHz, but for the ACELP algorithm the input signal is down-sampled to 12.8 kHz. For all coding modes in the AMR-WB, an asymmetric 384-sample long LP analysis window is used to generate a set of LP coefficients, in which 64 look-ahead samples are involved, resulting in a delay of 5ms.

The AMR-WB encoder includes Linear Prediction Coding (LPC) model in its coding process. The flow chart of the original Linear Prediction (LP) analysis model is shown in Figure 1. An input speech signal is received, and then the signal is pre-processed which includes the high-pass filtering and pre-emphasizing. After the pre-processing, short-term prediction, LP analysis is performed to calculate the LP coefficients $\{\alpha_i\}$ once per speech frame. The LP coefficients $\{\alpha_i\}$ are transformed to the Immittance Spectral Pair (ISP) domain for quantization and interpolation.

The LP analysis includes windowing, auto-correlation computation, and Levinson-Durbin algorithm. The principle of each part will be introduced in the following part with details.

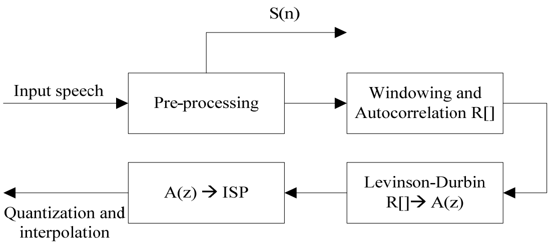

Figure 1. Flow chart of original algorithm

A. Windowing

LP analysis is performed once per frame using an asymmetric window. The window has its weight concentrated at the fourth sub-frame and it consists of two parts: the first part is a half of a Hamming window and the second part is a quarter of a cosine window [10]. The window is given by:

$$w(n)=\begin{cases}0.54-0.46\cos(\frac{2\pi n}{2L_1-1}), & n=0,1,...,L-1\\ \cos(\frac{2\pi(n-L_1)}{4L_2-1}) & n=L_1,L_1+1,...,L_1+L_2-1\end{cases} \quad (1)$$

Where the values L_1 =256 and L_2 =128 are used. The window shape is drawn in the Figure 2. A 30 ms asymmetric window and a look-ahead period of 5 ms are used in this windowing process.

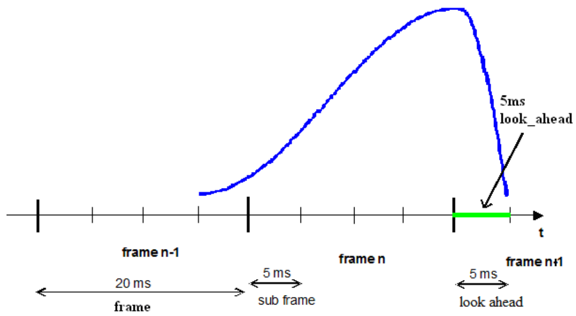

Figure 2. The window of AMR codec

B. Auto-correlation computation

Through windowing, windowed speech signal $s_w(n)$ is gotten. Where

$$s_w(n)=s(n)*w(n) \quad (2)$$

Then the autocorrelation function is gotten:

$$R(k)=\sum_{n=k}^{383}s_w(n)s_w(n-k) \qquad k=0,1,...,P \quad (3)$$

C. Levinson-Durbin algorithm

After the calculation of Auto-correlation computation, the Levinson-Durbin algorithm is used to calculate the Linear Prediction coefficients. The Levinson-Durbin algorithm will use the formula:

$$\sum_{i=1}^{P}\alpha_i R(|k-i|)=R(k) \qquad k=1,2,\cdots,P \quad (4)$$

Because AMR-WB uses the 16 order Linear Prediction analysis, we get Linear Prediction coefficients $\alpha_1,\alpha_2,...,\alpha_{16}$.

III. The proposed method

To skip the look-ahead samples and decrease the delay, an improved method is proposed and its flowchart is shown in Figure 3. An input speech signal is received, and then the signal is pre-processed. After the pre-processing, the autocorrelation method with new window is used to get LP coefficients $\{\alpha_i^{'}\}$. Then the LP coefficients $\{\alpha_i^{'}\}$ are used to produce the new samples S_{new} to replace the look-ahead samples. And then go on Linear Prediction (LP) analysis to calculate the LP coefficients $\{\alpha_i\}$. The LP coefficients $\{\alpha_i\}$ are transformed to the Immittance Spectral Pair (ISP) domain for quantization and interpolation.

Compared to the original method, the proposed method increased two steps.

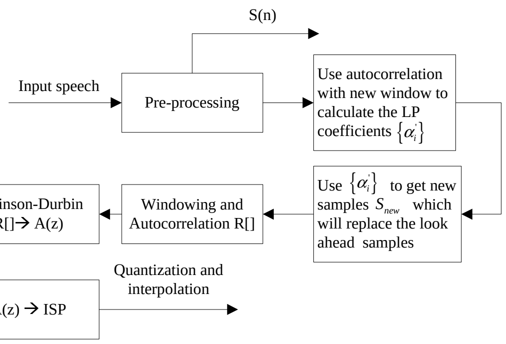

Figure 3. Flow chart of extrapolated method

A. change the window and calculate the new LP coefficients

After pre-processing, the autocorrelation uses a window with no look-ahead; as shown in Figure 4, the window is given by:

$$w^{'}(n)=\begin{cases}0.54-0.46\cos(\frac{2\pi n}{2L_1-1}), & n=0,1,...,L-1\\ \cos(\frac{2\pi(n-L_1)}{4L_2-1}) & n=L_1,L_1+1,...,L_1+L_2-1\end{cases} \quad (5)$$

Where the modified values L_1 =224 and L_2 =16 are used.

There are many kinds of window can be added here. After many tests, the window shown in Figure 5 is the best window for this method. The window contains only the samples of the current frame and the samples of the last two sub-frames of the previous frame.

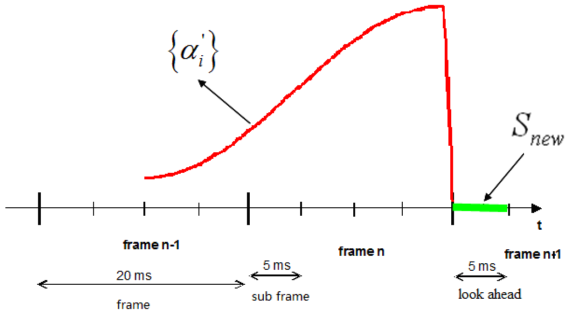

Figure 4. The window for producing new samples

Then windowed speech signal $s_w'(n)$ is gotten. Where

$$s_w'(n) = s(n) * w'(n) \tag{6}$$

Then the autocorrelation function is gotten:

$$R'(k) = \sum_{n=k}^{383} s_w'(n) s_w'(n-k) \qquad k = 0,1,\dots,P \tag{7}$$

Then LP coefficients $\{\alpha_i'\}$ is gotten from Levinson-Durbin algorithm

$$\sum_{i=1}^{P} \alpha_i' R'(|k-i|) = R'(k) \quad k = 1,2,\cdots,P \tag{8}$$

Here 16 orders Linear Prediction analysis is still used, therefore $P = 16$ and we get the new LP coefficients $\alpha_1', \alpha_2', \dots, \alpha_{16}'$.

B. use the new LP coefficients to predict the look-ahead samples

The LP coefficients $\{\alpha_i'\}$ are used to calculate the extrapolated samples s_{new} according to the following formula:

$$s_{new}(n) = \sum_{i=1}^{P} \alpha_i' \cdot s(n-i) \tag{9}$$

After the two steps, the original Linear Prediction analysis will go on. However the original look-ahead samples are replaced by s_{new}, and the new samples are used to windowing, Auto-correlation computation and Levinson-Durbin algorithm to calculate the linear prediction coefficients $\{\alpha_i\}$ for quantization and interpolation, which is shown in Figure 5.

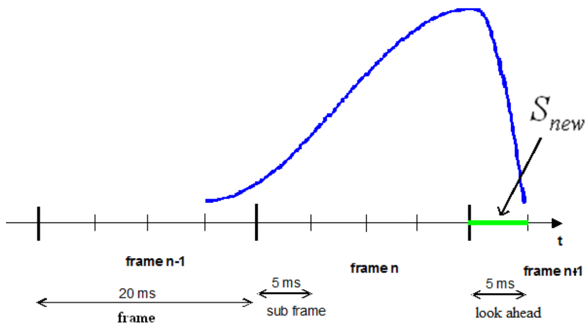

Figure 5. The window for calculating the LP coefficient

IV. Experimental results

To evaluate the quality of modified method, the objective and subjective evaluation methods are used.

A. The objective evaluation

Perceptual evaluation of speech quality (PESQ) is designed for the objective evaluation. The PESQ is done between the original AMR-WB speech codec and the modified method. Each PESQ is calculated through 20 samples with 10 female and 10 male Chinese samples. The rate chosen for the objective evaluation is 12.65kbps.

TABLE I. OBJECTIVE EVALUATION COMPARE BETWEEN ORIGINAL AND MODIFIED METHOD

AMR-WB 12.65kbps	Average PESQ	delay
Original method	3.701	25ms
Modified method	3.623	20ms

The PESQ results are shown in Table I. The PESQ of the modified method is only lower 0.8 than the original method, and the PESQ of the modified method is still higher than 3.5 which is the standard of communication quality. What's more, the delay of modified method is only 20ms which is lower 5ms than the original method. Therefore the result indicated that the reconstructed quality of the extrapolated method can reach that of the original AMR-WB codec, and at the same time, the modified method skip the look ahead period and decrease the delay.

B. The subjective evaluation

The A/B listening test method is used for the subjective quality evaluations. The A/B listening test method is chosen between in the modified method and the original method. The rate for AMR-WB is 12.65kbps. 10 samples composed of 5 female and 5 male Chinese samples is chosen, and 12 listening people composed of 6 males and 6 females are chosen.

The A/B listening test results are shown in Figure 6. Most of people choose the neutral, and the people chosen prefer the original method is only a little higher than the people chosen prefer modified method. It indicates that most of the people think the voice quality of modified method is near to the original method. Therefore the subject evaluation supports that the modified method can decrease delay without damage the voice quality.

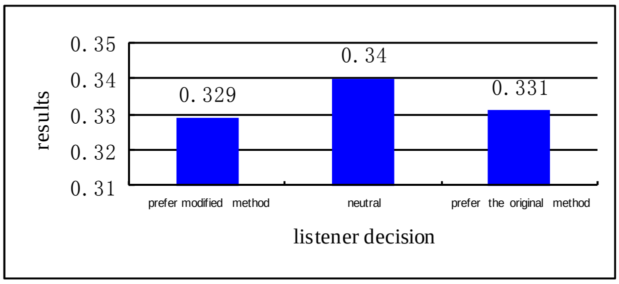

Figure 6. Average A/B results of 10 speech samples.

V. CONCLUSION

In this paper, a modified method is used to skip the look-ahead period. The modified method uses the autocorrelation algorithm to calculate the LP coefficients, and then the LP coefficients are employed to extrapolate new samples for replacing the look-ahead samples. Experiments result shows that the quality of this method is near to that of the original, at the same time the delay of the modified method is lower 5ms than the original method.

ACKNOWLEDGMENT

The work in this paper is supported by the cooperation project between Ericsson and Beijing Institute of Technology.

REFERENCES

[1] Bruno Bessette, Redwan Salami, and Roch Lefebvre, “The Adaptive Multirate Wideband Speech Codec (AMR-WB),” IEEE Trans. Speech and Audio Proc., vol.10, pp.620–636, 2002.

[2] Adaptive Multi-Rate Wideband (AMR-WB) Speech Codec; General Description, 3GPP TS 26.171, 2009.

[3] Yan Zhao, Yuyan Zhang, and Mang Jing, “AMR speech codec realization and optimization based on Tmscx55 DSP,” 2012 International Conference on Computer Science and Information Processing (CSIP), IEEE Press, Aug.2012, pp.120–123, doi: 10.1109/CSIP.2012.6308809.

[4] E. Ordentlich and Y. Shoham, “Low-delay code-excited linear predictive coding of wideband speech at 32 kbps,” in IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), IEEE Press, Apr.1991, pp. 9–12.doi: 10.1109/ICASSP.1991.150266.

[5] Adaptive Multi-Rate Wideband (AMR-WB) Speech Codec Transcoding functions, 3GPP TS 26.190, 2009.

[6] Adaptive Multi-Rate(AMR) Speech Codec Transcoding functions, 3GPP, TS 26.090, 2002.

[7] Makinen J, Bessette B, and Bruhn S. AMR-WB+: a new audio coding standard for 3rd generation mobile audio services[C]. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), IEEE Press, May.2005, pp.1109–1112, doi:10.1109/ICASSP.2005.1415603.

[8] Lan Juan, Huang Tie-Jun, Qu Jun-Hua, “A perception-based Scalable Encryption Model for AVS Audio,” in IEEE Int. Conf., Multimedia and Expo(ICME), IEEE Press, July.2007, pp.1778–1781, doi: 10.1109/ICME.2007.4285016.

[9] IETF. SILK Speech Codec, http://tools.ietf.org/html/draft-vos-silk-02,2010.

[10] Wai C. Chu, “Window Optimization in Linear Prediction Analysis,” IEEE Trans. Speech and Audio Proc., vol.11, pp.626–625, 2003.