An 800 bps vocoder based on Mixed Excitation Linear Prediction

Ye Li[1, a], Qiuyun Hao[1, b], Peng Zhang[1, c], Jingsai Jiang[1, d], XiaoFeng Ma[1, e], Yanhong Fan[1, f]

[1]Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center (National Supercomputer Center in Jinan), Jinan, 250014, China

[a]email: liye@sdas.org, [b]email:haoqy @sdas.org, [c]email: zhangp @sdas.org,

[d]email: jiangjs@sdas.org, [e]email:maxf @sdas.org, [f]email:fanyh @sdas.org,

Keywords: Vocoder; Mixed Excitation Linear Prediction; Channel Coding; Unequal Error Protection

Abstract. Based on the Mixed Excitation Linear Prediction (MELP) model, this paper presents a vocoder to obtain high-quality synthetic speech at 800 bps with 25% redundancy for channel coding. The vocoder parameters are designed and an unequal error protection method is proposed to improve the robustness over random error channel. Several channel coding schemes are compared and the optimal one is then selected. Test results show that the proposed speech coding algorithm could provide satisfactory speech quality and also have strong robustness for channel error.

Introduction

High quality speech coding algorithm with very low bit rate is one of the most important research topics. Speech coding algorithms with no more than 2.4kbps bit rate are applied widely in secure communications, satellite communications, radio communications, IP phone, etc. Linear prediction coding (LPC) is the most widely used algorithm for low-bit-rate speech coding, such as LPC-10 [1], Code Excited Linear Prediction (CELP) [2], and Mixed Excitation Linear Prediction (MELP) [3]. Due to the better quality, MELP vocoder has been selected as the U.S. Federal Standard for 2400 bps speech compression. 2.4 kbps to 300bps speech coding algorithms based on MELP have achieved great success and widespread applications [4-13]. However, the decoded speech quality is usually vulnerable to channel errors in such low bit rate speech coding algorithms. In this paper, the vocoder parameters are designed and an unequal error protection method is proposed to improve the robustness over random error channel. Several coding schemes are compared and then the optimal one is selected. Test results show that the proposed speech coding algorithm could provide satisfactory speech quality and also have strong robustness for channel error.

Principle of the speech coding algorithm

Principle of MELP

Mixed excitation and multi-band excitation are adopted in MELP algorithm which is based on the traditional LPC parametric model. In order to improve the quality of synthetic speech, several new techniques are adopted in the pitch extraction and excitation signal generation. Mixed excitation, aperiodic pulses, adaptive spectral enhancement, Fourier series and pulse dispersion are mainly included in MELP algorithm. Among them, mixed excitation, aperiodic pulses, adaptive spectral enhancement and Fourier series are introduced to further improve the quality of synthetic speech, while excitation signal and pulse spreading filter are used for sequential treatment of synthetic speech. The algorithm is improved based on the traditional MELP algorithm. The overall structures of the encoder and the decoder are shown in Figure 1 and Figure 2, respectively. Unlike traditional MELP algorithm, the Fourier amplitude parameters and aperiodic logo are abandoned, and only the line spectral frequency (LSF) parameters, band pass voicing coefficients (BPVC), pitch parameters and gain parameters are used for analysis, quantization, encoding and decoding.

Four parameters should be extracted for the speech coding algorithm based on MELP, which contains linear spectral frequency parameters (LSF), band pass voicing coefficients (BPVC), pith parameter and gain parameter, as shown in Figure 1. Gain parameters are calculated twice to once every sub-frame.

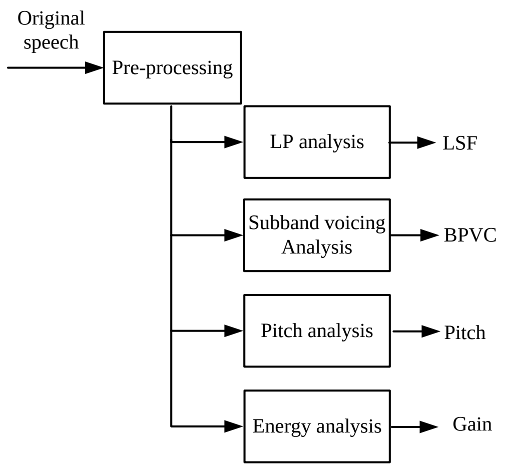


Fig.1 The overall encoder structure

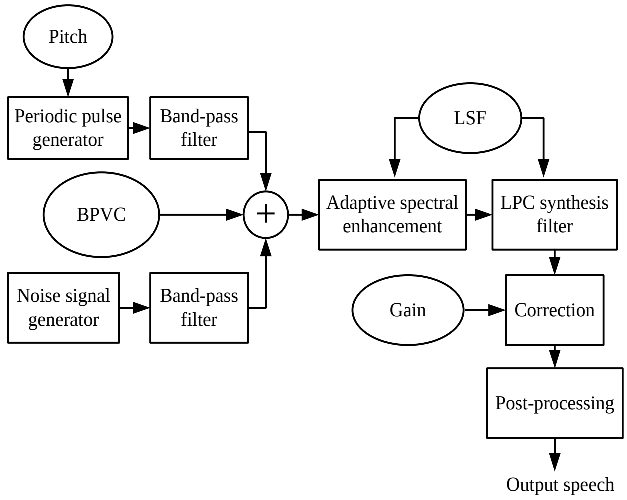


Fig. 2 The overall decoder structure

Bit Allocation Scheme

In the traditional MELP vocoder, the transmitted parameters are extracted every 22.5 ms frame (or 180 samples of speech at a sampling rate of 8 kHz). In the proposed vocoder, the transmitted parameters are extracted every 25 ms frame (or 200 samples of speech at a sampling rate of 8 kHz). A super frame structure of length 75ms comprising three consecutive frames is adopted in the proposed vocoder. Each super frame consists of 3 frames, for which the parameters are vector quantized, and the total number of quantized bits is 45. Such super frame structure is efficiently exploited by jointly vector quantization to reduce the inter frame redundancy. Thus, the source coding rate of the vocoder is 45bit/75ms=0.6kbps. Fourier amplitude parameters and aperiodic logo are mainly used to improve the naturalness of synthetic speech, but do not affect the definition and intelligibility. Considering the limit of bit rate, the Fourier amplitude parameters and aperiodic logo are abandoned since they have little effect on synthetic speech quality. The LSF parameters, BPVC, pitch parameters and gain parameters are retained since they have a greater effect on synthetic speech quality. The bit allocation scheme for the parameters is shown in Table 1.

Table 1. Bit allocation scheme for the parameters

Parameter	Bit Number
BPVC	3
pitch	9
gain	8
LSF	24
syn	1
Total	45

Quantization of Parameters

In the proposed speech coding algorithm, quantization scheme of LSF parameters, BPVC, pitch parameters and gain parameters is described as follows:

1) The BPVC parameters have a greater effect on the naturalness of synthetic speech. There are 5 sub-bands in a sub-frame, which is the same as in the MELP. The BPVC parameters of the three consecutive frames are grouped together into a 15 dimensional vector. According to the probability of occurrence of sub-band voicing, the unvoiced (U)/voiced (V) state of the three consecutive frames is represented in 8 modes, and the BPVC parameters are vector quantized

using 3 bits.

2) Since the human hearing is more sensitive to the quantization errors of pitch parameters, if the pitch parameters are not properly quantized, the tones of speech will produce distortions. The pitch parameters of the three consecutive frames (p_1, p_2, p_3) are grouped together into a 3 dimensional vector denoted by $\left[p_1,\ p_2,\ p_3\right]$. The three dimensional vector is then transformed to the logarithmic domain, which can be described as

$$P_{\lg n} = \left[\lg(p_1), \lg(p_2), \lg(p_3)\right] \tag{1}$$

The pitch parameters of the three consecutive frames are vector quantized using 9 bits with a codebook size of 512.

3) The gain parameters are vector quantized using 8 bits. There is only one gain parameter in each frame. The gain parameters of the three consecutive frames (g_1, g_2, g_3) are grouped together into a 3 dimensional vector denoted by $\left[g_1,\ g_2,\ g_3\right]$. The three dimensional vector is transformed to the logarithmic domain as described below:

$$G_{\lg n} = \left[\lg(g_1), \lg(g_2), \lg(g_3)\right] \tag{2}$$

The gain parameters of the three consecutive frames are vector quantized using 8 bits with a codebook size of 256.

4) The spectral envelope is mainly represented by the LSF parameters. The LSF parameters have a greater effect on the intelligibility of synthetic speech. Due to the slowly varying characters of LSF parameters, the inter frame has more redundant information. Therefore, the LSF parameters of the three consecutive frames are grouped together into a 30 dimensional vector. Vector quantization with inter-frame prediction is used to remove the redundancy, which means that the LSF parameter vector of the current super frame could be predicted with the LSF parameters of the final sub frame in the last super frame. The 30 dimensional vector is used to the prediction residual. Using multistage vector quantization with inter-stage prediction technology, the LSF parameters are vector quantized with four levels of codebook size of 128, 64, 64 and 32, corresponding to the quantization of LSF parameters using 7, 6, 6 and 5 bits respectively.

Channel Coding with Unequal Error Protection

According to the different importance for each parameter, unequal error protection (UEP) [14] can be used to provide different protection levels. In order to improve the transmission efficiency, the important code stream information is protected by using channel coding with strong error correcting ability, whereas the less important information is unprotected or protected by using channel coding with weak error correcting ability. UEP is introduced in this algorithm. Compared with Equal Error Protection (EEP) channel coding, the more important parameters or bits of the MELP vocoder are fully protected. Although the UEP channel coding can not reduce the bit error rate in random error channel, it can improve the quality of synthetic speech effectively.

The method of channel coding using UEP is described as follows:

1) Determine the important parameters or bits which can affect the quality of synthetic speech effectively. According to the speech coding algorithm of 600bps based on MELP and the results of experiment, the parameters or bits which need to be sufficiently protected are as follows:
   - The first groups of LSF parameters are the most important parameters, and then the importance of the remaining LSF parameters is attenuated in turn.
   - Gain parameters and pitch parameters are more important. Their errors will lead to sharp fluctuations of the signal to noise ratio and the tone.
2) The importance of coding bits is sorted according to the importance of the parameters. The main purpose of importance order is to facilitate the application of UEP channel coding.
3) Select channel coding scheme. The proposed vocoder is designed to provide a coded bit rate of

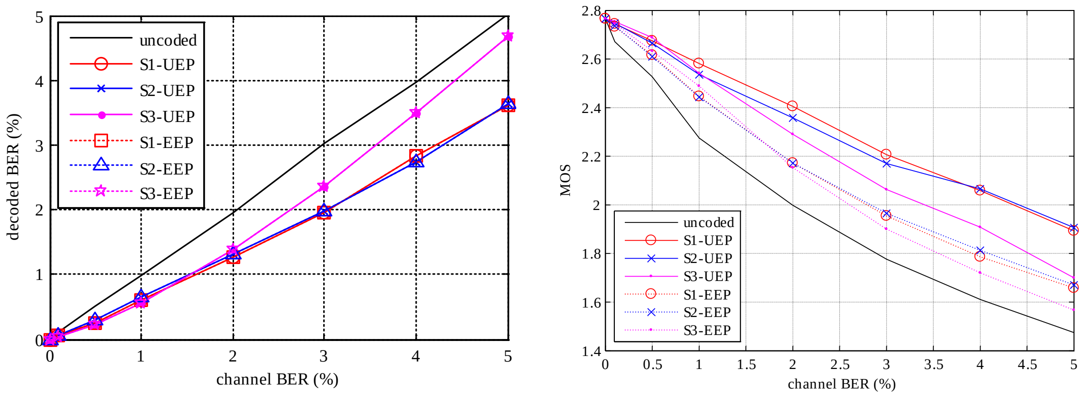
800 bps with 25% redundancy for channel coding. So for the channel coding scheme in each super frame, the total number of quantization bits is 45, and then it is encoded into 60 bits by using channel coding. In this algorithm, BCH code is adopted, which is a cyclic code that can correct multiple random errors. Its error correction ability is stronger, especially in short code length and medium long code length. Its performance is close to the theoretical value, and the structure is convenient and simple to implement. Since channel coding rate and source coding rate are limited, the key protection of bits and the available code group are considered to be very important.

Test Results and Conclusion

Considering the number of important bits and the available code groups, three schemes of unequal error protection channel coding are designed to implement the 800bps speech coding algorithm. The joint coding of 5 groups of BCH (7, 4) code is adopted in scheme S1 in which 20 important bits are protected and the remaining 25 bits are not protected. One group of BCH (31, 16) code is adopted in scheme S2 in which 16 important bits are protected and the remaining 29 bits are not protected. The joint coding of a group of BCH (15, 5) code and a group of BCH (31, 26) is adopted in scheme S3 in which 31 important bits are protected and the remaining 14 bits are not protected. According to the importance of the parameters, the parameters or bits which need to be sufficiently protected are as follows: the first groups of LSF parameters, the gain parameters, the pitch parameters, and the other important bits which are identified by experiments. The number of bits to be protected for each scheme is shown in Table 2.

Table 2. Number of bits to be protected for each scheme

Scheme		**Bit Number**			
		LSF	gain	pitch	BPVC
S1	5*BCH(7,4)	7	3	9	1
S2	BCH(31,16)	6	3	6	1
S3	BCH(15,5)+BCH(31,26)	13	8	9	1



(a) BER (b) MOS

Fig. 3 BER and MOS of each scheme at different channel error rates

The decoding bit error rate (BER) and the mean opinion score (MOS) of each scheme at different channel error rates are shown in Figure 3. According to the simulation results, the performance of the vocoder with channel coding is obviously better than that without channel coding. Compared with EEP channel coding, although the BERs of the two methods are equal, the MOS of the vocoder with UEP channel coding is significantly higher. The reason is that the importance of each bit is not distinguished in EEP channel coding scheme, in which only a limited number of bits are protected and some important bits are not sufficiently protected. In UEP channel coding scheme, the importance of each bit is sorted, and more protection ability is allocated for the important bits which have greater influence on the quality of synthetic speech. Therefore, the synthetic speech quality of the UEP scheme is better than that of the EEP scheme, no matter which channel coding method is used. The performance of different UEP channel coding schemes is different. When the channel BER is less than 5%, the performance of scheme S1 with UEP channel coding is optimal. Although the BER of S1 is not the lowest, this scheme provides satisfactory quality and strong error resilience since the important bits are strongly protected.

The objective MOS and subjective auditory tests show that the proposed 800 bps speech coding algorithm based on MELP could obtain high-quality synthetic speech when the channel BER is less than 5%.

Acknowledgement

This work was supported by the International Science and Technology Cooperation Program of China (Grant No. 2012DFR10500), by the Shandong Provincial Young and Middle-Aged Scientists Research Awards Fund of China (Grant No. BS2014DX019), and by the Youth Science Funds of Shandong Academy of Sciences (Grant No. 2014QN009).

References

[1] Thomas E. Tremain. The Government Standard Linear Predivtive Coding Algoritim: LPC-10. Speech Technology, 1982:40-49.

[2] Robert Fenichel, Dennis Bodson. Details to Assist in Implementation of Federal Standard 1016 CELP. NCS Technical Information Bulletin 92-1.

[3] A.V. McCree, T.P. Barnwell. A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding, IEEE Transactions on Speech and Audio Processing, 1995, 3(4):242-250.

[4] Wang T, Koishida K, Cuperman V, et al. A 1200 bps speech coder based on MELP. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Istanbul, Turkey, 2000:1375~1378.

[5] Wang T, Koishida K, Cuperman V et al. A 1200/2400 bps coding suite based on MELP. IEEE Workshop Proceedings on Speech Coding, Ibaraki, Japan, 2002: 90~92.

[6] NATO STANAG 4591 ANNEX M Ratification Draft 1. MELPe variation for 600 bit/s NATO narrow band voice coder. The Hague: NATO, 2003.

[7] Chong-White N.R., Cox R.V.. An intelligibility enhancement for the mixed excitation linear prediction speech coder. IEEE Transactions on Signal Processing Letters, 2003, 10(9): 263-266.

[8] de Lamare R.C. , aim A.. Strategies to improve the performance of very low bit rate speech coders and application to a variable rate 1.2 kb/s codec. IEE Proceedings: Vision, Image and Signal Processing, 2005, 152(1):74-86.

[9] Lahouti F. , Khandani A.K., Saleh A.. Robust Transmission of Multistage Vector Quantized Sources Over Noisy Communication Channels Applications to MELP Speech Codec. IEEE Transactions on Vehicular Technology , 2006, 55(6):1805-1811.

[10] Ye Li, Jingde Xu, Qinghua Li, Huijuan Cui, Kun Tang. A new distortion measure for parameter quantization based on MELP. International Journal of Image、Graphics and Signal Processing, 2010, 2:46-52.

[11] Ye Li, Jingsai Jiang, Qiang Guo. Reconstruction algorithm of Band Pass Voicing Coefficients based on mode transition and codebook mapping. The Sixth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2011, 2:970-972.

[12] Ye Li, Jingsai Jiang, Xiaofeng Ma . Research on Ultra-low-bit-rate speech coding algorithm for underwater speech communication. Applied Mechanics and Materials, 2011, 128-129:

247-250.

[13] Ye Li , Yan Hong Fan, Fei Yuan, Xiao Mei Xu. Research on an embedded ultra-low-bit-rate speech coding algorithm. Materials Science, Computer and Information Technology, 2014, 989-984:1951-1954.

[14] Lee-Fang Wei , AT&T Bell Lab., Holmdel, NJ, USA. Coded modulation with unequal error protection. IEEE Transactions on Communications, 1993, 41(10):1439 – 1449.