A 600bps Vocoder Algorithm Based on MELP
Lan ZHU and Qiang LI*
Chongqing Key Laboratory of Signal and Information Processing, Chongqing University of Posts and Telecommunications, Chongqing 400065, P. R. China
*Corresponding author

Keywords: Low-bit-rate speech coding, Multi-frame joint quantization, Vocoder.

Abstract. In order to meet the bandwidth requirements of low-bit-rate speech communication, there is an improvement of standard Mixed Excitation Linear Prediction (MELP) encoder and a 600bps MELP vocoder algorithm with higher quality is proposed. The algorithm forms a super frame by using 5 sub-frames. Joint quantization was carried out to remove redundancy information between intra-frames and inter-frames. We use the method of predictive split vector quantization to quantize the Line Spectral Frequency (LSF) parameters and the other parameters are jointly quantized. Compared to standard 2.4kbps MELP vocoder, the Mean Opinion Score (MOS) of the synthetic speech which is synthesized by the vocoder using the proposed algorithm is 2.54, and the synthetic speech quality is good.

Introduction
Low-rate speech coding technology is mainly used in the fields of secure communication, satellite communication, military communication and so on. In low-bit-rate speech coding standards, MELP uses mixed excitation signals as the inputs of a synthetic filter, and additional parameters is exploited on the basis of LPC in MELP to get more accurate excitation information. Therefore, MELP algorithm is mainly used in low-bit-rate vocoder. The 2.4kbps vocoder based on MELP algorithm became the USA federal standard in 1997[1]. Later, a 1.2/2.4kbps speech coder with noise preprocessor based on the MELP analysis algorithm identified by Department of Defense Digital Voice Processing Consortium (DDVPC) was selected as the North Atlantic Treaty Organization (NATO) standard STANAG4591[2].

Recent years, in order to reduce the bit rate of MELP vocoder, some scholars have implemented 600bps MELP vocoder. A 600bps low-bit-rate speech coding algorithm is proposed in [3]. This algorithm classifies the patterns of super frames, and it could be quantized for different super frame structures accurately. However, traditional multi-stage vector quantization methods were used to quantize LSF parameters, and the quality of synthetic speech still needs to be improved.

In this paper, we propose an improved 600bps MELP algorithm based on standard 2.4kbps MELP algorithm. The feature parameters in speeches are quantized by using the methods of multi-frame joint quantization, inter-frame prediction and linear interpolation. Particularly, we quantize the LSF parameters with the method of predictive split vector quantization, the quality of the synthesis speech can be prominently improved when compared with the quality using traditional multi-stage vector quantization. The experimental results show that this algorithm can reduce the bit rate effectively and high intelligibility of the synthesis speech can also be ensured.

Encode/Decode Structure
In 2.4kbps MELP vocoder standard, the speech is sampled at 8000HZ, the frame length is 22.5ms and every single frame is coded with 54 bits. During the process of encoding, parameters including LSF parameters, Pitch, Gain, Aperiodic Flag, Fourier magnitudes and Bandpass voicing strengths (BPVC) will be extracted and quantized. During the process of decoding, in order to improve the quality of synthetic speech, MELP vocoder will adopt the methods of mixed pulses, noise
excitation, adaptive enhancement, pulse scatter filtering and residual harmonic amplitude based on LPC vocoder.

In vocoder, the extraction accuracy of LSF parameters, Pitch, Gain and BPVC determine the quality of synthetic speech, while the other parameters, such as Fourier magnitudes, should be calculated based on other parameters' quantized results and it has little effect on the quality of synthesis speech. Therefore, in order to save quantization bits and reduce the complexity of the encoder algorithm, we quantize 4 parameters: LSF parameters, Pitch, Gain and BPVC.

**Parameter Quantization**

In this paper, we use the speech which is sampled at 8000HZ, and every 160 points as a frame, and 5 consecutive sub-frames form a super frame. We take super frame as a unit when quantizing parameters. In order to achieve the encoding rate of 600bps, we quantize and encode each super frame with 60bits. Specific quantization scheme is as follow.

**Division and Quantization of super frame structure**

When calculating unvoiced/voiced (U/V) states in a super frame, we use U to present the unvoiced frame, and use V to present the voiced frame. A super frame's 5 U/V states will be regarded as the super frame structure. By analyzing a large number of speech data, we count the occurrence probability of each super frame structure, and take 16 structures with the maximum probability of occurrence as the super frame states. The other 16 structures will be taken as the super frame state which is similar to it. Table 1 shows the division of the super frame states.

In the encoder, we need to quantize the parameters in different ways according to different super frame states. In the decoder, we decode parameters based on the super frame structure. Thus, we need transmit the 16 kinds of super frame state values. In this paper, the super frame states are scalar quantized with 4bits.

| State Value | Structure |
|-------------|-----------|
| 0 | UUUUU(UVUUU.UUUU.VUUUU.UVUVU.UVUVU) |
| 1 | VUUUU(VUUUV.VUVUV) |
| 2 | UUUUV(UVUUV.UUVUU) |
| 3 | VVUUU |
| 4 | VUUUV |
| 5 | VVUVU |
| 6 | UVVVV |
| 7 | VVVVU |
| 8 | VVVVV(VUVVV.VVUVV.VUVVU.VUVUV) |
| 9 | VVVUU(UVUVU.UVUVV) |
| 10 | UUVVV(UUVUV.UUVUU.UUVUV) |
| 11 | UUVVU |
| 12 | VVVUV |
| 13 | VVVUV |
| 14 | UUVVV |
| 15 | UUVVV |

**Quantization of LSF Parameters**

We use Predictive Split Vector Quantizer (PSVQ) to quantize 2 or 3 sub-frames' LSF parameters in a super frame, and the unquantized frames will be calculated by Lagrange interpolation method in the decoder by using the correlation between LSF parameters. The specific quantization steps are described as follows:

Firstly, we express LSF parameters as vector \( f = [f_0, f_1, \ldots, f_9] \).

Secondly, we remove the mean of LSF parameters, and get \( z(n) \):
Thirdly, we use 1-order moving average prediction method to find the predictive residual vector of current frame:

\[ r(n) = z(n) - p(n) \]  

(2)

Here, 1-order moving average prediction formula \( p(n) = 0.65 \hat{r}(n-1) \) is the predictive values of current frame, and \( \hat{r}(n-1) \) is the quantized LSF residuals of previous frame.

Finally, the calculated LSF residual vector \( r(n) \) will be quantized with the method of split vector quantization. If the frame is unvoiced, we quantize it with 9bits, divide \( r(n) \) into two vectors equally, and use 5bits and 4bits to quantize each vector correspondingly. If the frame is voiced, we quantize it with 16bits, divide \( r(n) \) into three vectors with 4, 3, 3elements, and use 7bits, 5bits and 4bits to quantize each vector correspondingly. The specific quantization scheme for LSF parameters is shown in table 2.

Table 2. The quantization of the LSF parameters.

| State Value | Structure | LSF parameters |
|-------------|-----------|----------------|
| 0 UUUUU     | 27        | quantize the 1st, 3rd, 5th frame with 9bit, 9bit, 9bit |
| 1 VUUUU     | 34        | quantize the 1st, 3rd, 5th frame with 16bit, 9bit, 9bit |
| 2 UUUUV     | 34        | quantize the 1st, 3rd, 5th frame with 9bit, 9bit, 16bit |
| 3 VVUUU     | 34        | quantize the 2nd, 3rd, 4th frame with 16bit, 9bit, 9bit |
| 4 VUUUV     | 34        | quantize the 2nd, 3rd, 5th frame with 9bit, 9bit, 16bit |
| 5 UVVUU     | 34        | quantize the 1st, 3rd, 5th frame with 9bit, 16bit, 9bit |
| 6 UUVVU     | 34        | quantize the 2nd, 3rd, 4th frame with 9bit, 9bit, 16bit |
| 7 UUUVV     | 34        | quantize the 2nd, 3rd, 4th frame with 9bit, 9bit, 16bit |
| 8 VVVVV     | 32        | quantize the 2nd, 4th frame with 16bit, 16bit |
| 9 VVVVU     | 32        | quantize the 1st, 3rd frame with 16bit, 16bit |
| 10 UVVVV    | 32        | quantize the 2nd, 5th frame with 16bit, 16bit |
| 11 UVVVU    | 32        | quantize the 1st, 5th frame with 16bit, 16bit |

Quantization of Pitch

First we take the log of every pitch. If there is no voiced frame in a super-frame, then the pitch will not be quantized. If there is one voiced frame in a super frame, we only need to quantize the pitch of the voiced frame with 7bits. If there is more than one voiced frame in a super frame, we form the pitch of a vector with 5 elements, and the vector will be quantized with 9 bits.
Quantization of Gain

We extract 2 gains in a sub-frame, form 10 gains of a vector with 10 elements in a super frame, and then the vector will be quantized with 8 bits.

Quantization of BPVC

BPVC is used to distinguish voicing characteristics of 5 sub-bands. We use 1 to present the unvoiced sub-band, and use 0 to present the voiced sub-band. The BPVC value of the lowest frequency sub-band is quantized with state value. Use the BPVC values of the rest 4 sub-bands to form a vector with 20 elements in a super frame, and then quantize the vector. If there is no voiced frame in the super frame, the BPVC will not be quantized. If there is one voiced frame in the super frame, we only need to quantize the vector with 4 bits. If there is more than one voiced frame in the super frame, we use 5 bits to quantize the vector.

Error Control

In this paper, the remaining bits are used to take Forward Error Correction (FEC). The FEC methods are as follows:

Step 1: In UUUUU mode, the 27 bits LSF parameter is protected by 7 (7, 4) Hamming codes, and the remaining information bit is 0.

Step 2: In VUUUU mode, the 4 bits state value is protected by a (7, 4) Hamming code.

Step 3: In UUUUV mode, the 4 bits state value is protected by a (7, 4) Hamming code.

Bit Allocation

In summary, we could get the quantization bit allocation scheme of 16 models for all parameters. Table 3 shows the bit allocation.

| State Value | Bit Allocation |
|-------------|----------------|
|             | State | Gain | LSF | Pitch | BPVC | Error Correction | Total |
| 0           | 4     | 8    | 27  | 0     | 0    | 21               | 60    |
| 1           | 4     | 8    | 34  | 7     | 4    | 3                | 60    |
| 2           | 4     | 8    | 34  | 7     | 4    | 3                | 60    |
| 3           | 4     | 8    | 34  | 9     | 5    | 0                | 60    |
| 4           | 4     | 8    | 34  | 9     | 5    | 0                | 60    |
| 5           | 4     | 8    | 34  | 9     | 5    | 0                | 60    |
| 6           | 4     | 8    | 34  | 9     | 5    | 0                | 60    |
| 7           | 4     | 8    | 34  | 9     | 5    | 0                | 60    |
| 8           | 4     | 8    | 32  | 11    | 5    | 0                | 60    |
| 9           | 4     | 8    | 32  | 11    | 5    | 0                | 60    |
| 10          | 4     | 8    | 32  | 11    | 5    | 0                | 60    |
| 11          | 4     | 8    | 32  | 11    | 5    | 0                | 60    |
| 12          | 4     | 8    | 32  | 11    | 5    | 0                | 60    |
| 13          | 4     | 8    | 32  | 11    | 5    | 0                | 60    |
| 14          | 4     | 8    | 32  | 11    | 5    | 0                | 60    |
| 15          | 4     | 8    | 32  | 11    | 5    | 0                | 60    |

Performance Testing

In this paper, the PESQ test software recommended in ITU-T Recommendations is used to test MOS of the algorithm. We select 16 test speeches (half males and half females) in P.862 standard as coding objects, and use the algorithm to implement a 600bps MELP vocoder, then encode and decode the test speeches to get synthetic speeches. We use PESQ software to test the MOS of synthetic speeches, and then compare the experimental results with the standard 2.4kbps MELP
vocoder and 1.2kbpsMELP vocoder studied in our laboratory, the test results are shown in Table 4. In the table, we use M stands for male, use F stands for female:

|       | 2.4kbps | 1.2kbps | 600bps |
|-------|---------|---------|--------|
|       | M       | F       | M       | F       | M       | F       |
| 1     | 3.49    | 3.40    | 3.39    | 3.37    | 2.66    | 2.84    |
| 2     | 3.34    | 3.45    | 3.22    | 3.24    | 2.75    | 2.71    |
| 3     | 3.19    | 3.34    | 3.16    | 3.09    | 2.64    | 2.78    |
| 4     | 3.61    | 3.43    | 2.90    | 2.73    | 2.51    | 2.38    |
| 5     | 3.64    | 3.33    | 2.83    | 2.95    | 2.29    | 2.37    |
| 6     | 3.46    | 3.22    | 2.78    | 3.24    | 2.44    | 2.43    |
| 7     | 3.54    | 3.12    | 2.68    | 2.98    | 2.56    | 2.18    |
| 8     | 3.39    | 3.35    | 3.12    | 3.12    | 2.63    | 2.58    |
| average | 3.46  | 3.33  | 3.01  | 3.09  | 2.56  | 2.53  |

As is shown in table 4, The Mean Opinion Score (MOS) of the synthetic speech which is synthesized by the vocoder using the proposed algorithm is 2.54, and the intelligibility of the speech is good.

**Conclusion**

In this paper, we make an improvement based on standard MELP vocoder and propose a high quality 600bps MELP vocoder algorithm. Through designing the frame structure and reasonably allocating quantization bits, the coding bit rate is reduced successfully. We also use different quantization methods for different super frame modes, and the qualities of the synthesized speech are improved. Compared with standard 2.4kbpsMELP vocoder and 1.2kbps MELP vocoder, the vocoder realized in this paper have high intelligibility, and can meet the quality requirements of voice communication.

**References**

[1] A. McCree, K. Truong, E. B. George, et al, A 2.4 kbit/s MELP coder candidate for the new US Federal Standard, IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 200-203, 1996.

[2] Tian Wang, Kazuhito Koishida, Vladimir Cuperman, Allen Gersho, John S. Collura, A 1200/2400 bps coding suite based on MELP, IEEE Workshop Proceedings, Tsukuba, Japan, pp. 90-92, 2002.

[3] Di Gao, Xiaojun Zhao, A 600bps MELP-based Speech Quantization Scheme for Underwater Acoustic Channels, IEEE International Conference on Computational and Information Sciences(ICCIS), pp. 1983-1986, 2013.

[4] J.C. Hardwick, J.S. Lim, A 4.8 kbps multi-band excitation speech coder, IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 374-377, 1988.

[5] Xiaofeng Ma, Ye Li, Jingsai Jiang, Peng Zhang, Yanhong Fan, Qiuyun Hao, 400bps High-Quality Speech Coding Algorithm, IEEE International Conference on Computer, Consumer and Control (IS3C), pp. 256-259, August 2016.

[5] Gwenael Guilmin, et al, New nato stanag narrow band voice coder at 600 bits/s, IEEE International Conference on Acoustics, Speech and Signal Processing, 2006.
[6] L.M. Supplee, R.P. Cohn, J.S. Collura, et al. MELP: the new federal standard at 2400 bps, IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 1591-1594, 1997.

[7] Q. Li, X. Hu, X. Xia, A design of 0.9 kb/s speech coding algorithm based on MELPe, IEEE International Congress on Image and Signal Processing, pp. 1580-1583, 2012.

[8] C. Wu, H. Jiang, B. Li, An improved MELP speech coder, IEEE International Conference on Information Technology and Computer Science, Vol. 2, pp. 130-133, 2009.

[9] T. Wang, K. Koishida, V. Cuperman, et al, A 1200 bps speech coder based on MELP, IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, pp. 1375-1378, 2000.

[10] C.L. Duta, L. Gheorghe, N. Tapus, Real time implementation of MELP speech compression algorithm using Blackfin processors, IEEE International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 250-255, 2015.

[11] Ye Li, Qiuyun Hao, Peng Zhang, Jingsai Jian, XiaoFeng Ma, Yanhong Fan, An 800 bps vocoder based on Mixed Excitation Linear Prediction, International Conference Mechanical Engineering, Industrial Electronics and Informatization, pp. 91-96, 2015.