Journal of Computer Science 6 (11): 1288-1292, 2010
ISSN 1549-3636

Multi-Pulse Based Code Excited Linear Predictive Speech Coder with Fine Granularity Scalability for Tonal Language

Suphattharachai Chomphan
Department of Electrical Engineering, Faculty of Engineering at Si Racha,
Kasetsart University, 199 M.6, Tungsukhla, Si Racha, Chonburi, 20230, Thailand

Abstract: Problem statement: The flexible bit-rate speech coder plays an important role in the modern speech communication. The MP-CELP speech coder which is a candidate of the MPEG4 natural speech coder supports a flexible and wide bit-rate range. However, a fine scalability had not been included. To support finer scalability of the coding rate, it had been studied in this study. **Approach:** In this study, based on the MP-CELP speech coding with HPDR technique, Fine Granularity Scalability was introduced by adjusting the amount of transmitted fixed excitation information. The FGS feature aim at changing the bit rate of the conventional coding more finely and more smoothly. **Results:** Through performance analysis and computer simulation, the quality of scalability of the MP-CELP coding was presented with an improvement from conventional scalable MP-CELP. The HPDR technique is also applied to the MP-CELP to use for tonal language, meanwhile it can support the core coding rate of 4.2, 5.5, 7.5 kbps and additional scaled bit rates. **Conclusion:** The core coder with high pitch delay resolution technique and adaptive codebook for tonal speech quality improvement has been conducted and the FGS brings about further efficient scalability.

Key words: Flexible bit-rate, speech coder, MP-CELP, fine granularity scalability, bit rate scalability, HPDR technique

INTRODUCTION

In the 3GPP CDMA systems, the EVRC speech coder performs very well with much more robustness than the older codec's. But for the bit rate range, it can support the range of 0.81-8.55 kbps. One candidate of the MPEG4 natural speech coder is MP-CELP which supports a more flexible and wider range of 5-29.5 kbps. This flexible coder employs the multi-pulse excitation which the number of pulses in fixed-entry codebook is selective for bit rate scalability and multiple bit rate functionality according to the MPEG-4 CELP speech coder requirements, (Nomura *et al*., 1998).

According to (Chompun *et al*., 2001a; 2003; Tan and Hussain, 2009; Al-Haddad *et al*., 2009; Haratyand El Ariss, 2007), a bit rate scalable tonal language speech coder based on a multi-pulse based code excited linear predictive coding is proposed. The coder provides the bit rate scalabilities which is effective in multimedia communications(Ozawa and Serizawa, 1998; Taumi *et al*., 1996; Ozawa *et al*., 1997). Moreover, this coder is improved for the tonal language speech by applying the high pitch delay resolutions to retain the tone information precision.

Generally, the bandwidth usage in a transmission channel cannot always be observed at the encoder side. As a consequence, improperly compressed data could be lost in case of the congested traffic. This problem can be solved by layer coding, i.e., a scalable bit stream consisting of a base layer followed by one or several enhancement layers. The base layer of which is the minimum requirement and has to be received by the decoder in order to maintain an acceptable quality of the decoded contain of the stream. The enhancement layers, on the other hand, are used to improve the media quality and it can be ignored one layer at a time.

In addition to layer coding, Fine Granularity Scalability (FGS) is a new approach, allowing the bitstream to be discarded with finer granularity instead of a whole layer. FGS provides the channel traffic supervisor a much easier and more flexible way to control the traffic. General audio and video coding algorithms with FGS have been adopted as part of the MPEG-4 international standard (Chen and I-Hsien, 2003). However, an FGS speech coding technique has not yet been standardized. The FGS algorithms used in MPEG-4 general audio and video share a common strategy, that the enhancement layers are distinguished by the different bit significance level at which a bit plane or bit array is sliced from the spectral residual.

Hence, the concept of FGS is applied to MP-CELP speech coding with HPDR technique. The excitation of

certain subframe can be generated just by extension of the previous one with little quality degradation. These ignored pulses information will be added back, one at a time, to the corresponding subframe by following proper modification. This can increase the reconstructed speech quality and this also implies that the granularity of the bitstream is in a single pulse basis.

MATERIALS AND METHODS

MP-CELP coder: In practical CELP-based speech coder, Linear Predictive Coding (LPC) model with only adaptive codebook always leaves errors between the synthesized speech and the original one. In a common process, the errors due to the imperfections f the model are compensated by stochastic, fixed codebook process. The stochastic process is often time implemented by fixed-code pulses which are added to the pitch path of the excitation. Then, the combined excitation vector is filtered through the LPC filter and the errors can therefore be minimized. Specifically, speech component generated by the fixed-code pulses is used to enhance the quality of synthesized speech in subframe basis as can be seen in Fig. 1 (Chen and I-Hsien, 2003).

The operation principle for bit rate scalable MP-CELP coder can be separated into 2 parts, MP-CELP core coder and bit rate scalable tool. The MP-CELP core coder achieves a high coding performance by introducing a multi-pulse vector quantization. This study uses at most 3 stages of the bit rate scalable tools according to the MPEG-4 CELP requirement. The bit rate scalable tool is connected to the core coder as illustrated in Fig. 2 (Chompun *et al*., 2001b). The bit rate scalable tool encodes the residual signal produced at the MP-CELP core coder utilizing the multi-pulse vector quantization. Adaptive pulse position control is employed to change the algebraic-structure codebook at each excitation-coding stage depending on the encoded multi-pulse excitation at the previous stage. The algebraic-structure codebook is adaptively controlled to inhibit the same pulse positions as those of the multi-pulse excitation in the MP-CELP core coder or the previous stage. The pulse positions are determined so that the perceptually weighted distortion between the residual signal and output signal from the scalable tool is minimized. The LP synthesis and perceptually weighted filters are commonly used for both the MP-CELP core coder and the scalable tool.

For this conventional coder, to support the functionality of multiple bit rates, the number of multi-pulse is chosen as 1, 5 and 10. The bit allocation is shown in Table 1.

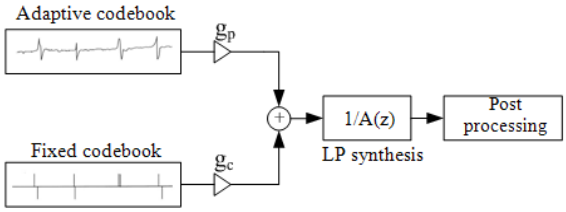

Fig. 1: Speech synthesizer

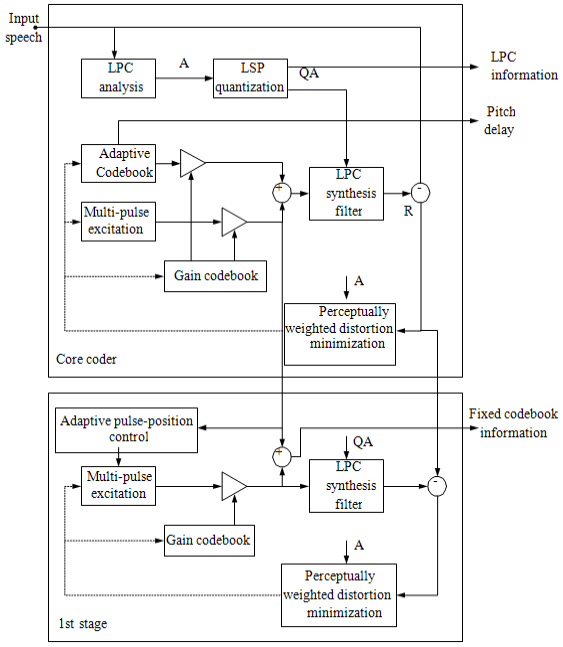

Fig. 2: Bit rate scalable MP-CELP coder

Table 1: Bit allocation of the conventional coder

	MP-CELP core coder			Bit rate scalable tool
Parameter	1-pulse core	5-pulse core	10-pulse core	(1 stage)
LSP	18	18	18	
Pitch delay	10	10	10	
Multi-pulse	7×2	20×2	40×2	4×2
Gain	7×2	7×2	7×2	
Total	56	82	122	8
Bit rate (bps)	5600	8200	12200	800

HPDR technique for tonal language speech: In Thai language, there are 5 different tones, mid (0), low (1), falling (2), high (3) and rising (4), whose characteristics are depicted in Fig. 3. Each graph represents the behavior of fundamental frequency (F0) in a period of syllable time where F0 is the inverse of pitch delay time. Though, F0 indicates the periodicity of voice. Investigating the difference between Thai male and Thai female F0 behaviors, Thai female F0 change rate is almost all more than Thai male F0's, (Thathong *et al*., 2000).

This is why the Thai female speech quality encoded by CS-ACELP coder is lower than the Thai male speech quality (Chompun *et al*., 2000; Laflamme *et al*., 1991). Hence, detecting F0 with high precision yields the improvement of the tonal language speech quality.

Since pitch delay (or F0) significantly involves in tone of tonal language, this study proposes an improvement of the bit rate scalable MP-CELP coder by applying the High Pitch Delay Resolutions (HPDR) technique to the pitch analysis of the core coder. The HPDR at pitch fraction of 1/2, 1/3 and 1/4 is adopted to the pitch analysis, consequently, it causes the increments of bit rate as 200, 400 and 400 bps respectively.

The HPDR technique is done by including the pitch fraction analysis within the conventional pitch analysis which finds the optimum fraction around the prior pitch delay integer of the conventional pitch analysis. In order to find the adaptive excitation for the proposed technique, the FIR filter based on a Hamming windowed sin(x)/x function truncated at ±11 and padded with zeros at ±12 is adopted to weight the excitation in the pitch fraction analysis (Chompun *et al*., 2001b).

Bit rate reduction method: In some cases, the channel bandwidth could be too small and this would limit the transmission of coded bitstream. Properly adapting the coding algorithm might be the solution but not an efficient way. According to (Chen and I-Hsien, 2003) the number of fixed-code pulses, which occupies a big percentage of the total bit rate, can be cut in half by ignoring those pulses in the odd numbered subframes. Since this procedure drops the pulses information in odd numbered subframes, the pulses searching procedure of these subframes is skipped in this bit-rate reduction approach. In other words, pulses information in odd numbered subframe will not be received or take part in synthesis. The total excitation of odd numbered subframe is constructed just by the excitation of the previous subframe.

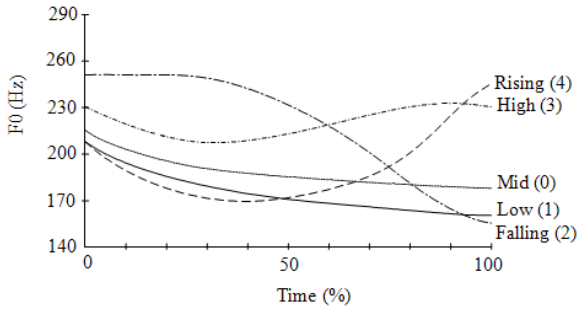

Fig. 3: F0 characteristic of 5 tones in Thai

Base on the (Chen and I-Hsien, 2003), FGS can be achieved by searching the discarded pulses of odd numbered subframes and delicately adding them back to the bitstream. In other words, the information bits associated with the fixed-code pulses of odd numbered subframes can be viewed as the enhancement layer of the bitstream. The following detail describes the modification involved in realizing this concept.

MP-CELP Based FGS: The enhancement layer of an FGS bitstream is allowed to be discarded as a whole or by part depending on the transmission environment. Placing the odd numbered subframe pulses in the enhancement layer implies that the number of those pulses received by the decoder is unknown at the encoder side. The purpose of the analysis-by-synthesis approach, by embedding a decoder in the encoding process, is for the encoder to foresee the exact speech decoded by the decoder on the other end of the transmission line. If the encoder has no knowledge about the number of odd numbered subframe pulses actually used by the decoder, it would have no base for constructing the best parameters to be sent to the decoder. This phenomenon could jeopardize the analysis-by-synthesis method used in the standard coder.

One way to minimize this problem is to assume the worst case of the receiving condition, i.e., always assume that the decoder receives none of the information bits from the enhancement layer. To be more precise in terms of implementation, the excitation vector and the memory states (of the LPC filtering) passed over from an odd numbered subframe to the next even-numbered subframe have to be constructed without any information from the odd numbered subframe pulses.

Table 2: Coding rate of the proposed coder (bps)

		Scaled Rate of 1-pulse-based FGS					
Core rate		FGS1	FGS2	FGS3	FGS4	FGS5	FGS6
No odd	4600	5000	5400	5800	6200	6600	7000
subframe	5900	6300	6700	7100	7500	7900	8300
	7900	8300	8700	9100	9500	9900	10300
With odd	6000	6400	6800	7200	7600	8000	8400
subframe	8600	9000	9400	9800	10200	10600	11000
	12600	13000	13400	13800	14200	14600	15000

Table 3: Conventional coding rate of MP-CELP with HPDR technique (bps)

	Conventional scaled rate		
Core rate	Layer 1	Layer 2	Layer 3
6000	6800	7600	8400
8600	9400	10200	11000
12600	13400	14200	15000

The odd numbered subframe pulses are still searched and generated, however they are purely used for extra quality enhancement of that subframe and are never recycled in the future subframes. If the encoder is allowed to recycle any of the odd numbered subframe pulses which are not received by the decoder, then, the coded parameters selected for the next subframe might not be the right choice for the decoder and an error would occur. The supporting coding rates are shown in Table 2, while the conventional coding rates are shown in Table 3.

RESULTS

The coding quality of the proposed coder was evaluated subjectively and objectively. The effectiveness of the spontaneous FGS and FGS ignoring those pulses in the odd numbered subframes was evaluated using both average MOS scores and segPSNR (segmental Power Signal to Noise Ratio) in dB (Chompun *et al*., 2001a). First, the comparison tests between the spontaneous FGS and conventional scalable MP-CELP were conducted and shown in graphs of Fig. 4 and 5. Figure 4 shows the speech quality of them in MOS scores, while Fig. 5 shows the speech quality in segPSNR.

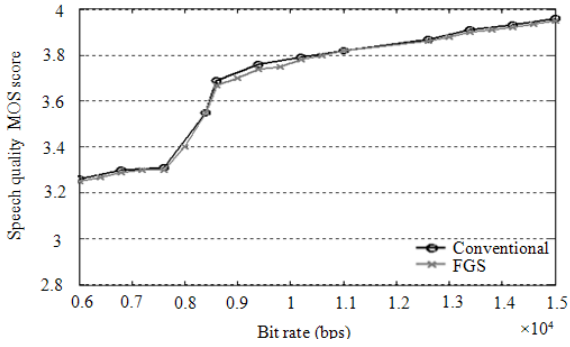

Fig. 4: Speech Quality of FGS and conventional scalable MP-CELP in MOS

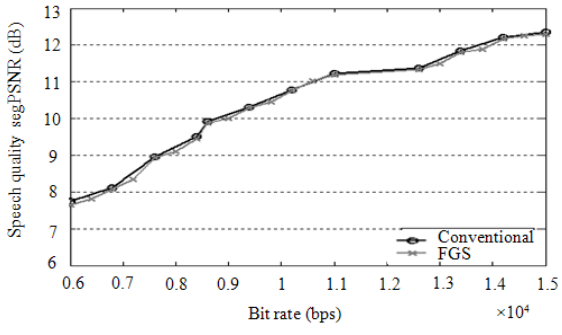

Fig. 5: Speech Quality of FGS and conventional scalable MP-CELP in segPSNR

DISCUSSION

The results from Fig. 4 and 5 shows that the speech quality of the coder with FGS is little lower than that of the conventional coder for all levels of bit rate, on the other hand, the FGS can support twice as many bit rates as the conventional coder. This indicates that the proposed FGS brings about better scalability attribute.

Subsequently, the comparison tests between the FGS ignoring those pulses in the odd numbered subframes and conventional scalable MP-CELP were conducted and shown in graphs of Fig. 6 and 7. Figure 6 shows the speech quality of them in MOS scores, while Fig. 7 shows the speech quality in segPSNR. The results from Fig. 6 and 7 shows that the speech quality of the coder with FGS is noticeable lower than that of the conventional coder for all levels of bit rate due to the suppression of pitch information in odd numbered subframes. However, the proposed FGS can provide much more bit rates than the conventional coder and serve the system at very low bit rates.

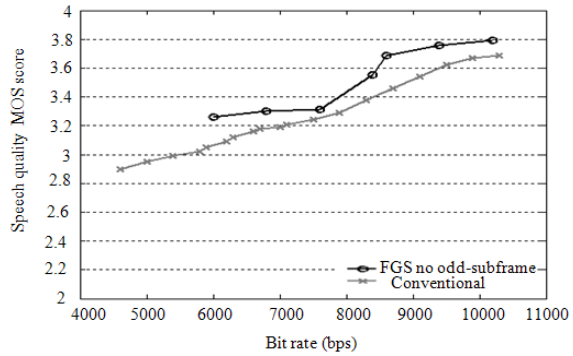

Fig. 6: Speech Quality of FGS ignoring those pulses in the odd numbered subframes and conventional scalable MP-CELP in MOS

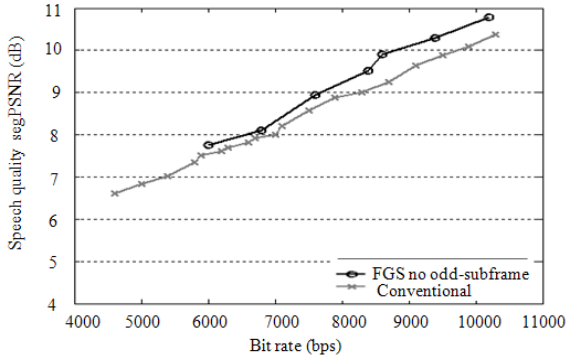

Fig. 7: Speech Quality of FGS ignoring those pulses in the odd numbered subframes and conventional scalable MP-CELP in segPSNR

CONCLUSION

A modification of bit rate scalable tonal language speech coder has been proposed. This coder consists of a MP-CELP core coder with and the FGS bit rate scalable tools. The core coder employs high pitch delay resolutions with adaptive codebook for tonal speech quality improvement. The FGS brings about the efficient scalability. The results show that coder provides more supporting levels of coding rate, meanwhile, the quality of the proposed coder is little sacrificed for FGS attribute.

REFERENCES

Al-Haddad, S.A.R., S.A. Samad, A. Hussain, K.A. Ishak and A.O.A. Noor, 2009. Robust speech recognition using fusion techniques and adaptive filtering. Am. J. Applied Sci., 6: 290-295. http://www.scipub.org/fulltext/ajas/ajas62290-295.pdf

Chompun, S., S. Jitapunkul, D. Tancharoen and T. Srithanasan, 2000. Thai speech compression using CS-ACELP coder based on ITU G.729 standard. Proceeding of the 4th Symposium on Natural Language Processing, May 10-12, NECTEC, Chiangmai, Thailand, pp: 1-5. http://daisy.ee.eng.chula.ac.th/~d1oatty/oat_files/group_files/paper/SNLP2000_Supattarachai_final.pdf

Chompun, S., S. Jitapunkul and D. Tancharoen, 2001a. Novel technique for tonal language speech compression based on a bit rate scalable MP-CELP coder. Proceeding of the IEEE International Conference on Information Technology: Coding and Computing, Apr. 2001, IEEE Xplore Press, Las Vegas, USA., pp: 461-464. DOI: 10.1109/ITCC.2001.918839

Chompun, S., S. Jitapunkul, D. Tancharoen and P. Kittipanya-ngam, 2001b. HPDR Technique for tonal language speech compression based on MP-CELP coder with multiple bitrates and bitrate scalabilities. Proceeding of the World Multiconference on Systemics, Cybernetics and Informatics, July, 22-25, IIIC, Orlando, Florida, USA., pp: 1-4.

Chompun, S., Y. Yothinsumpun, D. Tancharoen and S. Jitapunkul, 2003. Performance evaluation of multi-pulse code-excited linear-predictive coder with high pitch delay resolutions technique over additive white Gaussian noise and Rayleigh fading channels. Proceeding of the International Conference on Information and Communication Technologies, Apr. 8-10, Assumption University, Thailand, pp: 71-75. http://cat.inist.fr/?aModele=afficheN&cpsidt=18597026

Chen, F.C. and L. I-Hsien, 2003. CELP Based speech coding with fine granularity scalability. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, Apr. 6-10, IEEE Xplore Press, USA., pp: 145-148. DOI: 10.1109/ICASSP.2003.1202315

Haraty, R.A. and O. El Ariss, 2007. CASRA+: A colloquial Arabic speech recognition application. Am. J. Applied Sci., 4: 23-32. http://www.scipub.org/fulltext/ajas/ajas4123-32.pdf

Laflamme, C., J.P. Adoul, R. Salami, S. Morissette and P. Mabilleau, 1991. 16 kbps wideband speech coding technique based on algebraic CELP. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 14-17, IEEE Xplore Press, Toronto, Ont., Canada, pp: 13-16. DOI: 10.1109/ICASSP.1991.150267

Nomura, T., M. Iwadare, M. Serizawa and K. Ozawa, 1998. A bitrate and bandwidth scalable CELP coder. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 12-15, IEEE Xplore Press, Seattle, USA., pp: 341-344. DOI: 10.1109/ICASSP.1998.674437

Ozawa, K. and M. Serizawa, 1998. High quality multi-pulse based CELP speech coding at 6.4 kb/s and its subjective evaluation. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 12-15, IEEE Xplore Press, Seattle, USA., pp: 153-156. DOI: 10.1109/ICASSP.1998.674390

Ozawa, K., T. Nomura, M. Serizawa, 1997. MP-CELP speech coding based on multi-pulse vector quantization and fast search. Elect. Commun. Jap. Part III: Fund. Elect. Sci., 80: 55-63. DOI: 10.1002/(SICI)1520-6440(199711)80:11<55::AID-ECJC6>3.0.CO;2-R

Tan, T.S. and S. Hussain, 2009. Corpus design for Malay corpus-based speech synthesis system. Am. J. Applied Sci., 6: 696-702. http://www.scipub.org/fulltext/ajas/ajas64696-702.pdf

Taumi, S., K. Ozawa, T. Nomura and M. Serizawa, 1996. Low-delay CELP with multi-pulse VQ and fast search for GSM EFR. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 7-10, IEEE Xplore Press, Atlanta, USA., pp: 562-565. DOI: 10.1109/ICASSP.1996.541158

Thathong, U., S. Jitapunkul, V. Ahkuputra, E. Maneenoi and B. Thampanitchawong, 2000. Classification of Thai consonants naming using Thai tone. Proceeding of the International Conference on Spoken Language Processing, Oct. 16-20, ISCA, Beijing, China, pp: 47-50. http://www.isca-speech.org/archive/icslp_2000/i00_3047.html