Journal of Computer Science 7 (1): 65-69, 2011
ISSN 1549-3636

High Pitch Delay Resolution Technique for Tonal Language Speech Coding Based on Multi-Pulse Based Code Excited Linear Prediction Algorithm

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: In spontaneous speech communication, speech coding is an important process that should be taken into account, since the quality of coded speech depends on the efficiency of the speech coding algorithm. As for tonal language which tone plays important role not only on the naturalness and also the intelligibility of the speech, tone must be treated appropriately. **Approach:** This study proposes a modification of flexible Multi-Pulse based Code Excited Linear Predictive (MP-CELP) coder with multiple bitrates and bitrate scalabilities for tonal language speech in the multimedia applications. The coder consists of a core coder and bitrate scalable tools. The High Pitch Delay Resolutions (HPDR) are applied to the adaptive codebook of core coder for tonal language speech quality improvement. The bitrate scalable tool employs multi-stage excitation coding based on an embedded-coding approach. The multi-pulse excitation codebook at each stage is adaptively produced depending on the selected excitation signal at the previous stage. **Results:** The experimental results show that the speech quality of the proposed coder is improved above the speech quality of the conventional coder without pitch-resolution adaptation. **Conclusion:** From the study, it is a strong evidence to further apply the proposed technique in the speech coding systems or other speech processing technologies.

Key words: High Pitch Delay Resolutions (HPDR), Multi-Pulse based Code Excited Linear Predictive (MP-CELP), speech coding, bitrate scalability, multiple bitrates, Linear Prediction (LP), Line Spectrum Pairs (LSP), Tone (T)

INTRODUCTION

Nowadays the digital mobile communications are widely developed. The speech, audio, images, video or data information can be transmitted through wire or wireless network channels (Jabrane *et al*., 2007). Simultaneously, the number of users to access these networks increases rapidly. Consequently, channel capacity has to be increased, signal compression aims to perform this (Chompun *et al*., 2000). Since the multimedia applications such as videophone and videoconference on ATM and Internet are widely used, the high quality speech coders are highly demanded. These kinds of applications require special considerations for packet loss. To overcome this problem, it is to realize a scalable coder where the synthesized speech signal can be decoded from the received packets, which contain only a part of the whole encoded bitstream. One of standardization activities for such areas is undergoing at the MPEG-4 (Nomura *et al*., 1998; Chomphan, 2010b; Sen, 2005).

In 1995, Conjugate-Structure Algebraic Code Excited Linear Predictive (CS-ACELP) coding was developed and standardized as ITU G.729 speech coding at 8 kbps. Later, MP-CELP coder has been proposed to be a scalable coder around this bitrate. This flexible coder employs the multi-pulse excitation which the number of pulses in fixed-entry codebook is selective for bitrate scalability and multiple bitrate functionality according to the MPEG-4 CELP speech coder requirements, see e.g., (Nomura *et al*., 1998; Chomphan, 2010b; Melesse and Hanley, 2005).

In MP-CELP, amplitudes or signs for multi-pulse excitation are simultaneously vector quantized. To improve speech quality for background noise conditions, the adaptive pulse location restriction method are applied (Ozawa and Serizawa, 1998). This coder operates at various bitrates ranging from 4-12 kbps utilizing the flexibility in multi-pulse excitation coding (Chomphan, 2010a; Ghaderi, *et al*., 2005).

As for tonal language, such as Thai, a syllable is composed of consonants, vowels and tone (Wutiwiwatchai and Furui, 2007). The smallest structure of sounds or syllables in Thai is composed of one vowel unit or one diphthong, one, two or three consonants and a tone. The structure can be represented as illustrated in Fig. 1. Ci is initial consonant, Cf is final consonant, V is vowel and T is tone.

$$C_i(C_i)\ V(V)\overset{T}{C_f}$$

Fig. 1: Thai syllable structure

The significant difference between tonal and toneless language is Tone (T). In tonal language, the words of different tones yield their distinguished meaning. By using the standard speech coder such as CS-ACELP with tonal language, it showed the degraded speech quality when compared to those of toneless language. The reason is that the tone information precision is not enough for tonal language, e.g., (Chompun *et al*., 2000; Wutiwiwatchai and Furui, 2007).

This study proposes a bitrate scalable tonal language speech coder based on a multi-pulse based code excited linear predictive coding (Taumi *et al*., 1996; Ozawa *et al*., 1996; Sen *et al*., 2005; Ghrare *et al*., 2009). The proposed coder provides the bitrate scalabilities which is effective in multimedia communications. Moreover, this coder is improved for the tonal language speech by applying the high pitch delay resolutions to retain the tone information precision.

MATERIALS AND METHODS

Bitrate scalable MP-CELP coder: The operation principle for bitrate scalable MP-CELP coder can be separated into 2 parts, MP-CsELP core coder and bitrate scalable tool.

MP-CELP core coder: The MP-CELP core coder achieves a high coding performance by introducing a multi-pulse vector quantization as depicted in Fig. 2 (Taumi *et al*., 1996; Ozawa *et al*., 1996). The input speech of 10 ms frame is processed through Linear Prediction (LP) and pitch analysis. The LP coefficients are quantized in the Line Spectrum Pairs (LSP) domain. The pitch delay is encoded by using an adaptive codebook. The residual signal for LP and the pitch analysis is encoded by the multi-pulse excitation scheme. The multi-pulse excitation signal is composed of several non-zero pulses. The pulse positions are restricted in the algebraic-structure codebook and determined by an analysis-by-synthesis approach, e.g., (Laflamme *et al*., 1991; Chomphan, 2010a). The pulse signs and positions are encoded, while the gains for pitch predictor and the multi-pulse excitation are normalized by the frame energy and encoded.

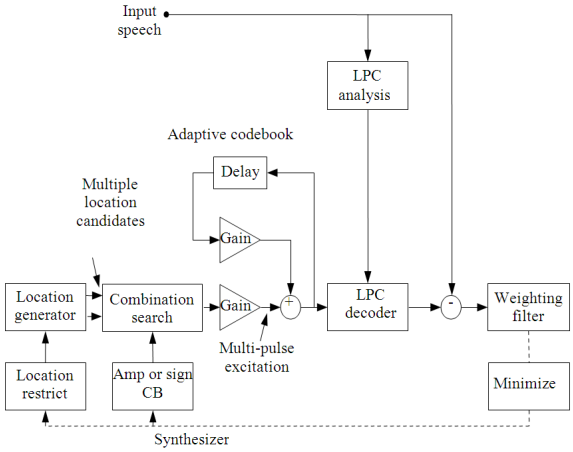

Fig. 2: MP-CELP core coder

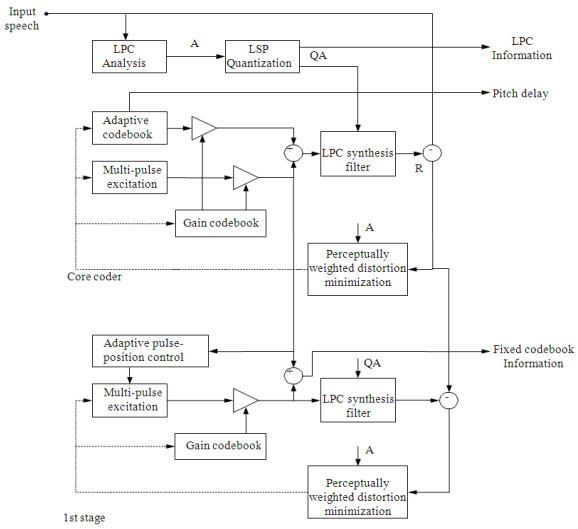

Fig. 3: One-stage bitrate scalable MP-CELP coder

Bitrate scalable tool: This study uses at most 3 stages of the bitrate scalable tools according to the MPEG-4 CELP requirement. The bitrate scalable tool is connected to the core coder as illustrated in Fig. 3 (Chomphan, 2010a; 2010b). The bitrate 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 bitrates, the number of multi-pulse is chosen as 1, 5 and 10. The bit allocation is shown in Table 1. As for bitrate scalable tool, each stage increases the bitrate of 800 bps. Though, as for 1 multi-pulse, the total bitrate are 5600, 6400, 7200 and 8000 bps respectively. As for 5 multi-pulses, the total bitrate are 8200, 9000, 9800 and 10600 bps respectively. And as for 10 multi-pulses, the total bitrate are 12200, 13000, 13800 and 14600 bps respectively.

Tonal language speech coder : 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. 4 (Chompun *et al*., 2000; Wutiwiwatchai and Furui, 2007). 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, see e.g., (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). Hence, detecting f0 with high precision yields the improvement of the tonal language speech quality.

Table 1: Bit allocation for the conventional coder

Parameter	MP-CELP core coder	Bitrate scalable tool (1 stage)
LSP	18	
Pitch delay	10	
Multi-pulse	7×2, 50×2, 40×2	4×2
Gain	7×2	
Total	56	8
Bitrate (bps)	5600, 8200, 12200	800

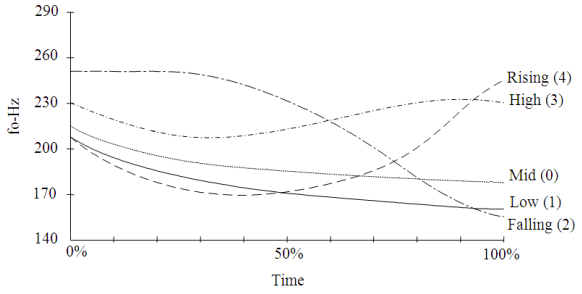

Fig. 4: f0 characteristic of 5 tones in Thai

Since pitch delay (or f0) significantly involves in tone of tonal language, this study proposes an improvement of the bitrate 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 bitrate 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.

RESULTS

The coding quality of the proposed coder was evaluated subjectively and objectively by using 36 tested sentences from 16 men and 16 women, some of them were shown in Table 2.

Table 2: Thai tested sentences (examples) 0, 1, 2, 3, 4 at each word represent tone of Thai

Order	Tested sentences
1	เขา แห่ นาค เวียน รอบ โบสถ์
	khaw4 hx:1 na:k2 wi:an0 r@:p2 bo:t1
2	คน ทำ บาป อวด ตัว ว่า เก่ง
	khon0 tham0 ba:p1 ?u:at1 tu:a0 wa:2 keng1
3	คำ ว่า เตียบ แปร ว่า ตะ ลุ่ม
	kham0 wa:2 ti:ap1 plx:0 wa:2 ta1 lum2
4	พวก นั้น โดน ปรับ ราย ตัว
	phu:ak2 nan3 do:n0 prap1 ra:j0 tu:a0
5	เขา เป็น ญาติ อำ ภา
	khaw4 pen0 ja:t2 ?am0 pha:0
6	น้อง จะ เอา ว่าว อัน นั้น
	n@:ng3 ca1 ?aw0 waw2 ?an0 nan3
7	เขา ยากอ สัก ลาย เสือ ที่ แขน
	khaw4 ja:k1 sak1 la:j0 sv:a4 thi:2 khx:n4

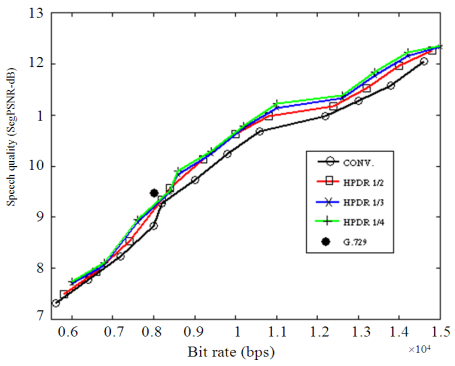

Fig. 5: Objective speech quality (SegSNR)

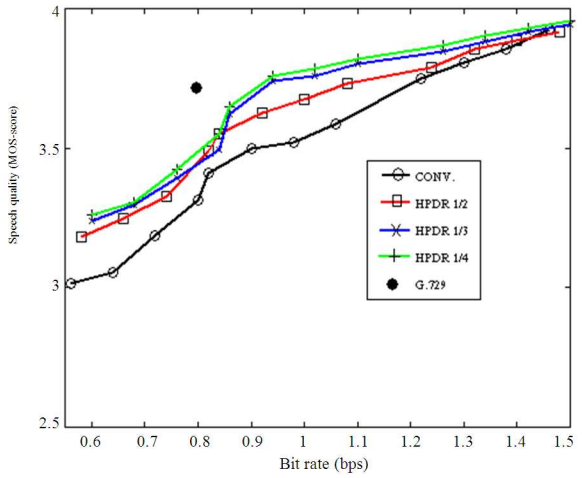

Fig. 6: Subjective speech quality (MOS Score)

The effectiveness of the high pitch delay resolutions applied to the conventional coder was evaluated using average segmental SNRs and MOS scores. Comparison tests of each grouped bitrate were conducted and shown in graphs of Fig. 5 and 6.

DISCUSSION

For the objective test (SegSNR), graphs in Fig. 5 showed that both male and female speech quality, every grouped bitrates, the HPDR at pitch fraction of 1/4 gave the maximum value. The order of speech quality from the best to the worst was 1/4's, 1/3's, 1/2's and conventional's respectively. For the subjective test (MOS score), the results in Fig. 6 were corresponding to those of the objective test.

The experimental results showed that the higher resolution, the more speech quality. This indicates that the proposed HPDR technique brings about better pitch precision which causes the improvement of the coding quality for tonal language.

CONCLUSION

A modification of bitrate scalable tonal language speech coder has been proposed. This coder consists of a MP-CELP core coder and the bitrate scalable tools. The high pitch delay resolutions are applied to adaptive codebook of core coder for tonal speech quality improvement. The results show that the coding quality of the proposed coder is better than the conventional coder for Thai language.

ACKNOWLEDGEMENT

The researchers are grateful to Digital Signal Processing Research Laboratory, Chulalongkorn University for providing the facility and technical support to this research and Kasetsart University at Si Racha campus for the research scholarship through the board of research.

REFERENCES

Chomphan, 2010a. Multi-pulse based code excited linear predictive speech coder with fine granularity scalability for tonal language. J. Comput. Sci., 6: 1288-1292. DOI: 10.3844/jcssp.2010.1288.1292

Chomphan, 2010b. Performance evaluation of multi-pulse based code excited linear predictive speech coder with bitrate scalable tool over additive white gaussian noise and rayleigh fading channels. J. Comput. Sci., 6: 1433-1437. DOI: 10.3844/jcssp.2010.1433.1437

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

Ghaderi, S.F., M.A. Azadeh and S. Bamdad, 2005. Analyzing the electricity consumption using experimental design technique. Am. J. Applied Sci., 2: 1464-1470. DOI: 10.3844/.2005.1464.1470

Ghrare, S.E., M.A. Mohd. Ali, K. Jumari and M. Ismail, 2009. An efficient low complexity lossless coding algorithm for medical images. Am. J. Applied Sci., 6: 1502-1508. DOI: 10.3844/.2009.1502.1508

Jabrane, Y., R. Iqdour, B.A. Es Said and N. Naja, 2007. MAI cancellation in DS/CDMA using a new approach on WDS. Am. J. Applied Sci., 4: 736-740. DOI: 10.3844/ajassp.2007.736.740

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

Melesse, A.M. and R.S. Hanley, 2005. Energy and carbon flux coupling: multi-ecosystem comparisons using artificial neural network. Am. J. Applied Sci., 2: 491-495. DOI: 10.3844/.2005.491.495

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, IEEE, Seattle, USA, May 12-15, 1998, pp: 341-344. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=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, Seattle, USA., pp: 529-532. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=674390

Ozawa, K., T. Nomura and M. Serizawa, 1996. MP-CELP speech coding based on multi-pulse vector quantization and fast search. IEICE Trans., 79: 1655-1663. DOI: 10.1002/(SICI)1520-6440(199711)80:11<55::AID-ECJC6>3.0.CO;2-R

Sen, M.D.L., 2005. Asymptotic hyperstability of dynamic systems with point delays. Am. J. Applied Sci., 2: 1279-1282. DOI: 10.3844/.2005.1279.1282

Sen, M.D.L., J.L. Malaina, A. Gallego and J.C. Soto, 2005. Stability of non-neutral and neutral dynamic switched systems subject to internal delays. Am. J. Applied Sci., 2: 1481-1490. DOI: 10.3844/.2005.1481.1490

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 and V. Ahkuputra, 2000. Classification of Thai consonants naming using Thai tone. Proceeding of the International Conference on Spoken Language Processing, Beijing, China, pp: 47-50. http://www.isca-speech.org/archive/icslp_2000/i00_3047.html

Wutiwiwatchai, C. and S. Furui, 2007. Thai speech processing technology: A review. Speech Commun., 49: 8-27. DOI: 10.1016/j.specom.2006.10.004