

Tonal Language Speech Compression Based on a Bitrate Scalable Multi-Pulse Based Code Excited Linear Prediction Coder

Suphattharachai Chomphan

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

Abstract: Problem statement: Speech compression is an important issue in the modern digital speech communication. The functionality of bitrates scalability also plays significant role, since the capacity of communication system varies all the time. When considering tonal speech, such as Thai, tone plays important role on the naturalness and the intelligibility of the speech, it must be treated appropriately. Therefore these issues are taken into account in this study. **Approach:** This study proposes a modification of flexible Multi-Pulse based Code Excited Linear Predictive (MP-CELP) coder with bitrates scalabilities for tonal language speech in the multimedia applications. The coder consists of a core coder and bitrates scalable tools. The high pitch delay resolutions are applied to the adaptive codebook of core coder for tonal language speech quality improvement. The bitrates 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, the proposed approach is able to improve the speech compression quality for tonal language and the functionality of bitrates scalability is also developed.

Key words: Thai speech, Multi-Pulse based Code Excited Linear Predictive (MP-CELP), speech compression, bitrates scalability, Conjugate-Structure Algebraic Code Excited Linear Predictive (CS-ACELP), Tone (T), bitrates scalable, High Pitch Delay Resolutions (HPDR), MOS scores

INTRODUCTION

In the present days, the digital communications are widely developed (Jabrane *et al.*, 2007). The audio, images, video or data information can be transmitted through wire or wireless network channels. 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 codes 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).

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 bitrates. This flexible coder employs the multi-pulse excitation which the number of pulses in fixed-entry codebook is selective for bitrates scalability and multiple bitrates functionality according to the MPEG-4 CELP speech coder requirements, (Nomura *et al.*, 1998; Chomphan, 2010b; Rahim and Islam, 2009).

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, 2010b; Singh *et al.*, 2007).

As for tonal language, such as Thai, a syllable is composed of consonants, vowels and tone (Wutiwiwatchai and Furui, 2007).



Fig. 1: Thai syllable structure

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.

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, (Chompun *et al.*, 2000; Wutiw WATCHAI and Furui, 2007).

This study proposes a bitrates scalable tonal language speech coder based on a multi-pulse based code excited linear predictive coding (Taumi *et al.*, 1996; Ozawa *et al.*, 1996; Al-Haddad *et al.*, 2009). The proposed coder provides the bitrates 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

Bitrates scalable MP-CELP coder: The operation principle for bitrates scalable MP-CELP coder can be separated into 2 parts, MP-CELP core coder and bitrates 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, (Laflamme *et al.*, 1991; Chomphan, 2010a).

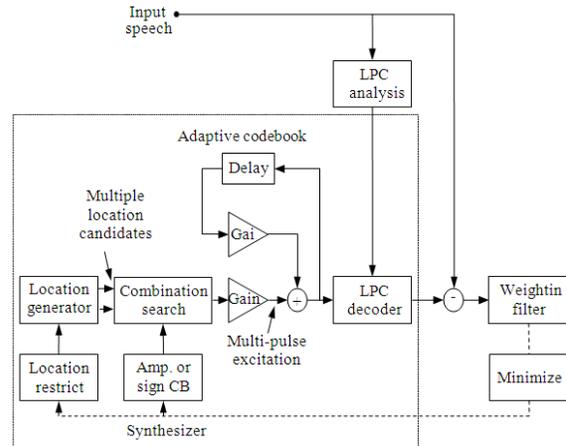


Fig. 2: MP-CELP core coder

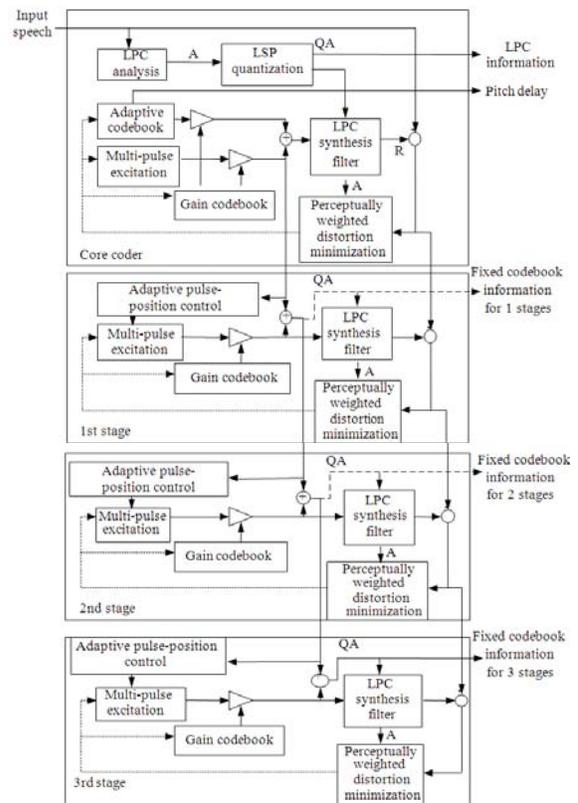


Fig. 3: Three-stage bitrate scalable MP-CELP coder

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.

Bitrates scalable tool: This study uses at most 3 stages of the bitrates scalable tools according to the MPEG-4CELP requirement as illustrated in Fig. 3 (Chomphan,

2010b). The bitrates 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, the bit allocation is shown in Table 1. The bitrates of core coder is 5600 bps. As for bitrates scalable tool, each stage increases the bitrates of 800 bps. Though, the 1, 2, 3 stages of scalability operate at the total bitrates of 6400, 7200 and 8000 bps respectively.

Tonal language speech coder: In Thai speech, there are 5 different tones, mid(0), low(1), falling(2), high(3) and rising(4), whose characteristics are depicted in Fig. 4 (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, (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. Since pitch delay (or f0) significantly involves in tone of tonal language, this study proposes an improvement of the bitrates 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 bitrates 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.

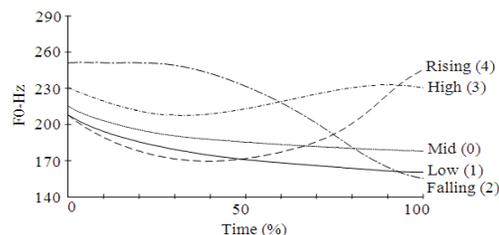


Fig. 4: The f0 characteristic of 5 tones in Thai

Table 1: Bit allocation for the conventional coder

Parameter	MP-CELP core coder	Bitrates scalable tool (1 stage)
LSP	18	
Pitch delay	14	
Multi-Pulse	5×2	4×2
Gain	7×2	
Total	56	8
Bitrates (bps)	5600	800

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

Order	Tested sentences
1	เขา แท้ นาค เวียน รอบ โปสธ khaw4 hx:1 na:k2 wi:an0 r@:p2 bo:t1
2	คน ทำ มาป ลอด 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 อ่า ภา khaw4 pen0 ja:t2 ?am0 pha:0
6	น้อง จะ เอา ว่า ว่า อัน นั้น n@:ng3 ca1? aw0 waw2? an0 nan3
7	เขา ออยาก ลัก ลาย เสื่อ ที่ แขน Khaw4 ja:k1 sak1 la:j0 sv:a4 thi:2 khx:n4

Table 3: Objective speech quality (SegSNR)

Stages of scalability	none	1	2	3	
Grouped bitrates (pbs)	5600+	6400+	7200+	8000+	
Male	Conv.	7.45	7.81	8.31	9.06
	HPDR.1 2 ⁻¹	7.61	7.93	8.79	9.45
	HPDR.1 3 ⁻¹	7.70	8.01	8.87	9.59
	HPDR.1 4 ⁻¹	7.86	8.04	8.90	9.61
Female	Conv.	7.26	7.66	8.18	8.69
	HPDR.1 2 ⁻¹	7.35	7.88	8.21	9.23
	HPDR.1 3 ⁻¹	7.61	8.01	8.82	9.38
	HPDR.1 4 ⁻¹	7.70	8.06	8.85	9.41

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 10 men and 10 women, some of them were shown in Table 2.

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 bitrates were conducted and shown in Table 3 and 4.

Table 4: Subjective speech quality (MOS Score)

Stages of scalability	none	1	2	3	
Grouped bitrates (pbs)	5600+	6400+	7200+	8000+	
Male	Conv.	3.12	3.18	3.28	3.35
	HPDR.1/2	3.18	3.22	3.30	3.36
	HPDR.1/3	3.21	3.30	3.32	3.38
	HPDR.1/4	3.22	3.33	3.34	3.39
Female	Conv.	3.09	3.11	3.20	3.31
	HPDR.1/2	3.13	3.16	3.24	3.33
	HPDR.1/3	3.16	3.23	3.27	3.34
	HPDR.1/4	3.19	3.28	3.30	3.36

DISCUSSION

For the objective test (SegSNR), the results 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 were corresponding to those of the objective test.

From the experimental results, the more high resolution be used the more speech quality be obtained. This indicates that the proposed HPDR technique brings the better pitch precision which causes an improvement of the coding quality for tonal language.

CONCLUSION

In this study, a modification of bitrates scalable tonal language speech compression has been proposed. The compression algorithm consists of a MP-CELP core coder and the bitrates 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. The proposed approach can support a number of coding rates.

ACKNOWLEDGEMENT

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

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. DOI: 10.3844/ajassp.2009.290.295

Chomphan, S., 2010a. 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: 1438-1442. DOI: 10.3844/jcssp.2010.1433.1437

Chomphan, S., 2010b. 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

Chompun, S., S. Jitapunkul, D. Tancharoen and T. Srihanasan, 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

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. Mabillean, 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*, IEEE, May 12-15, 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*, IEEE, May 12-15, Seattle, WA, USA., pp: 153-156. DOI: 10.1109/ICASSP.1998.674390

Ozawa, K., T. Nomura and M. Serizawa, 1996. MP-CELP speech coding based on multipulse vector quantization and fast search. *Elec. Electronics Eng.*, 80: 55-63. DOI: 10.1002/(SICI)1520-6440(199711)80:11<55::AID-ECJC6>3.0.CO;2-R

Rahim, N.A. and Z. Islam, 2009. Field programmable gate array-based pulse-width modulation for single phase active power filter. *Am. J. Applied Sci.*, 6: 1742-1747. DOI: 10.3844/ajassp.2009.1742.1747

- Singh, M.S.J., S.I.S. Hassan and M.F. Ain, 2007. Rainfall attenuation and rainfall rate measurements in malaysia comparison with prediction models. *Am. J. Applied Sci.*, 4: 5-7. DOI: 10.3844/ajassp.2007.5.7
- 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. Ahkputra, 2000. Classification of Thai consonant naming using Thai tone. *Proceeding of the International Conference on Spoken Language Processing*, Oct. 16-20. Beijing, China, pp: 47-50
- Wutiwivatchai, C. and S. Furui, 2007. Thai speech processing technology: A review. *Speech Commun.*, 49: 8-27. DOI: 10.1016/j.specom.2006.10.004