American Journal of Applied Sciences 9 (3): 350-353, 2012
ISSN 1546-9239

Analytical Study of High Pitch Delay Resolution Technique for Tonal Speech Coding

[1]Suphattharachai Chomphan and [2]Chutarat Chompunth
[1]Department of Electrical Engineering, Faculty of Engineering at Si Racha, Kasetsart University, 199 M.6, Tungsukhla, Si Racha, Chonburi, 20230, Thailand
[2]School of Social and Environmental Development, National Institute of Development Administration, 118 M.3, Serithai Road, Klong-Chan, Bangkapi, Bangkok, 10240, Thailand

Abstract: Problem statement: In tonal-language speech, since tone plays important role not only on the naturalness and also the intelligibility of the speech, it must be treated appropriately in a speech coder algorithm. **Approach:** This study proposes an analytical study of the technique of High Pitch Delay Resolutions (HPDR) applied to the adaptive codebook of core coder of Multi-Pulse based Code Excited Linear Predictive (MP-CELP) coder. **Results:** The experimental results show that the speech quality of the MP-CELP speech coder with HPDR technique is improved above the speech quality of the conventional coder. An optimum resolution of pitch delay is also presented. **Conclusion:** From the analytical study, it has been found that the proposed technique can improve the speech coding quality.

Key words: High Pitch Delay Resolutions (HPDR), Multi-Pulse based Code Excited Linear Predictive (MP-CELP), speech coding, speech compression, tone

INTRODUCTION

Speech coding is an important process in the present digital mobile communications. The number of users to access the communication networks increases rapidly. As a result, channel capacity has to be increased, in which the speech compression or coding aims to perform this (Chompun *et al*., 2000).

MP-CELP coder has been proposed to be a scalable coder for Moving Picture Expert Group-4 (MPEG-4) speech coder standards at low bit rate. 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 (Nomura *et al*., 1998; Chomphan, 2010b). 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).

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

The important difference between tonal and toneless language is the existence of Tone. In tonal language, the words with 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 presents a technique of high pitch delay resolutions or HPDR for a bitrate scalable tonal language speech coder based on a multi-pulse based code excited linear predictive coding. It aims at preserving the tone information precision. The experimental results show the efficiency of the HPDR technique with different resolutions.

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

Fig. 1: Thai syllable structure

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

MATERIALS AND METHODS

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*., 1997). The input speech of 10 m sec 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.

HPDR technique: An important parameter of MP-CELP speech coder is pitch delay which is inversely propotional with the fundamental frequency. Basically, the fundamental frequency contour determines the characteristics of tones. In summary, to treat the tonal characteristics precisely, the pitch delay should be analyzed correctly and precisely.

Instead of using the pitch delay with integer number, we apply the pitch fraction in the order of one second, one third and one fourth.

Applying the pitch fraction at one second (1/2), the additional bit information that must be included in the output bitstream is 200 bps. Meanwhile, applying the pitch fraction at one third (1/3) and one fourth (1/4), the additional bit information that must be included in the output bitstream is 400 bps.

Pitch fraction analysis: It is done by considering the cross correlation of the target signal and the excitation signal in the previous stage or in the buffer memory. Applying the pitch fraction at one second (1/2), the optimal pitch fraction corresponds to the fraction that maximizes the cross correlation function in the following Eq. 1:

$$R(k)_t = \sum_{i=0}^{2} R(k-i)b(t+i.2) + \sum_{i=0}^{2} R(k+1+i)b(2-t+i.2), \quad t = 0,1 \tag{1}$$

where, b(n) is the weighting function generated from the function since (n) bounded by the following Hammimg window function, Eq. 2 and Fig. 3:

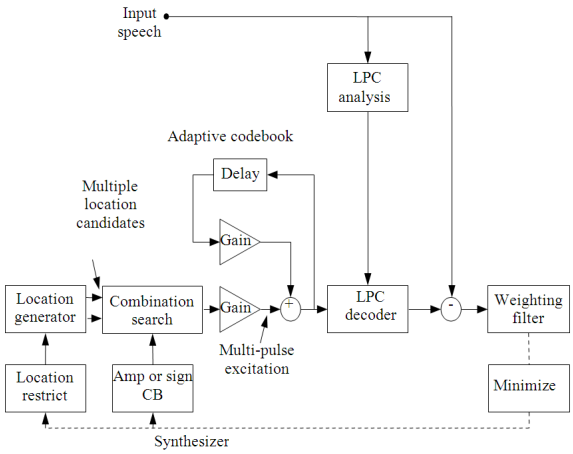

Fig. 2: MP-CELP core coder

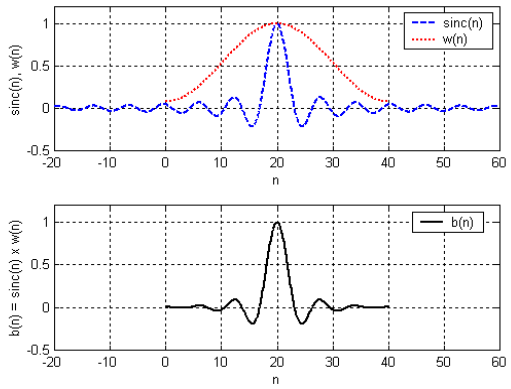

Fig. 3: Weighting function b (n) generated from sinc(n) multiplied with Hamming window w(n)

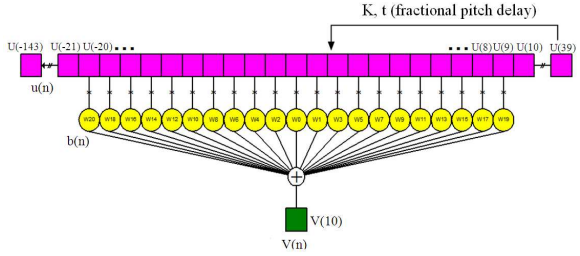

Fig. 4: Adaptive excitation signal generation using a 21-sample weighting function

$$w(n) = \begin{cases} 0.54 - 0.46\cos(2\pi n / m), & 0 \le n < m+1 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

In the maximization process of cross correlation function, the optimal pitch delay (k) and pitch fraction

(t) are used to obtain the optimal excitation signal v (n) by going back to excitation signal u (n) with corresponding time distance in the buffer memory as shown in Eq. 3 and Fig. 4:

$$v(n) = \sum_{i=0}^{W} u(n-k-i)b(t+i.2) + \sum_{i=0}^{W} u(n-k+1+i)b(2-t+i.2) \quad (3)$$

To apply the pitch fraction at one third (1/3), Eq. 4 is used instead of Eq. 1 and Eq. 5 is used instead of Eq. 3. Finally, to apply the pitch fraction at one fourth (1/4), Eq. 6 is used instead of Eq. 1 and 7 are used instead of Eq. 3:

$$R(k)_t = \sum_{i=0}^{3} R(k-i)b(t+i.3) + \sum_{i=0}^{3} R(k+1+i)b(3-t+i.3), \quad t = 0,1,2 \quad (4)$$

$$v(n) = \sum_{i=0}^{W} u(n-k-i)b(t+i.3) + \sum_{i=0}^{W} u(n-k+1+i)b(3-t+i.3) \quad (5)$$

$$R(k)_t = \sum_{i=0}^{4} R(k-i)b(t+i.4) + \sum_{i=0}^{4} R(k+1+i)b(4-t+i.4), \quad t = 0,1,2,3 \quad (6)$$

$$v(n) = \sum_{i=0}^{W} u(n-k-i)b(t+i.4) + \sum_{i=0}^{W} u(n-k+1+i)b(4-t+i.4) \quad (7)$$

RESULTS

The coding quality of the MP-CELP speech coder with HPDR technique was evaluated subjectively by using 36 tested sentences from 16 men and 16 women. The Hamming window width is varied from 5-37 samples. The sign "-" in Table 1 denotes the conventional coder without HPDR (1/2) technique.

DISCUSSION

From Table 1, it has been seen that the coding quality increases when the Hamming window width is increased.

Table 1: Subjective speech quality (MOS score)

HPDR at various hamming window width		MOS score: Core rate of 5600 bps	MOS score: Core rate of 8200 bps	MOS score: Core rate of 12200 bps
Hamming	-	3.02	3.41	3.75
window width	5	3.06	3.44	3.73
	9	3.11	3.48	3.76
	13	3.14	3.51	3.76
	17	3.16	3.52	3.78
	21	3.18	3.55	3.79
	25	3.19	3.55	3.80
	29	3.18	3.56	3.78
	33	3.17	3.53	3.79
	37	3.19	3.57	3.78

It is noted that the coder with HPDR technique gives the higher score than that of the conventional coder for all core bitrates. Moreover the coding quality of MP-CELP speech coder at the core bitrate of 12200 bps gives the highest score, while the coding quality of MP-CELP speech coder at the core bitrate of 5600 bps gives the lowest score at the same Hamming window width.

CONCLUSION

This study presents the HPDR technique to improve the coding quality for tonal language such as Thai. This core coder is based on MP-CELP speech coder. 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.

REFERENCES

Chomphan, S., 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, S., 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: 1438-1442. 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. Proceedings of the 4th Symposium on Natural Language Processing, (SNLP' 00), NECTEC, Chiangmai, Thailand, pp: 1-5.

Laflamme, C., J.P. Adoul, R. Salami, S. Morissette and P. Mabilleau, 1991. 16 kbps wideband speech coding technique based on algebraic CELP. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Apr. 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. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE Xplore Press, Seattle, USA, May 12-15, IEEE Xplore Press, Seattle, WA , USA., pp: 341-344. DOI: 10.1109/ICASSP.1998.674437

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

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

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

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