Journal of Computer Science 6 (12): 1438-1442, 2010
ISSN 1549-3636

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

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 the modern speech communication technology, the speech coding with bitrate scalability was needed. However, various types of noises in the communication channels cause damages in the transmitted information especially speech data. Tonal-language speech was also affected by this situation. **Approach:** Based on the high pitch delay resolution technique, the MP-CELP speech coding was proposed over the environment with CDMA AWGN and Rayleigh fading channels for tonal language. The proposed coder supports multiple bitrates and also has the functionality of bitrate scalability. **Results:** Through performance analysis and computer simulation, the quality of the proposed coding was presented with an improvement from conventional scalable MP-CELP in the specific-noise environments. The HPDR technique was applied to the MP-CELP to use for tonal language, meanwhile it can support the core coding rate of 5.6, 8.2, 12.2 kbps and additional scaled bitrates. **Conclusion:** By applying the high pitch delay resolution technique with the MP-CELP speech coding, we can improve the quality of tonal encoded speech. Moreover, the coding quality of the proposed coder was better than the conventional coder for Thai language over both AWGN channel and Rayleigh fading channel.

Key words: Tonal speech, MP-CELP speech coding, AWGN, Rayleigh fading channels, Bitrate scalability, multiple bitrates

INTRODUCTION

Nowadays the digital communications are widely developed. The audio, images, video or data information can be transmitted pass 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; Chompun, 2004). 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).

As for the 3GPP CDMA systems, the EVRC speech coder performs very well with much more robustness than the older codec’s (Jabrane *et al*., 2007). 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 bitrate scalability and multiple bitrate functionality according to the MPEG-4 CELP speech coder requirements (Nomura *et al*., 1998).

The performance of MP-CELP with High Pitch Delay Resolutions (HPDR) technique is presented in this study by examining in time varying channels. Result for Rayleigh flat fading channels is compared to the AWGN channel in the context of cellular communication environment (Adetunde and Seidu, 2008; Vlasie and Rousseau, 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 (Nomura *et al*., 1998).

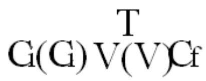

Fig. 1: Thai syllable structure

As for tonal language, such as Thai, a syllable is composed of consonants, vowels and tone (Chompun *et al*., 2001a; 2001b). 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; Chompun *et al*., 2001a; 2001b).

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*., 1997). 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-CELP core coder and bitrate scalable tool.

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 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.

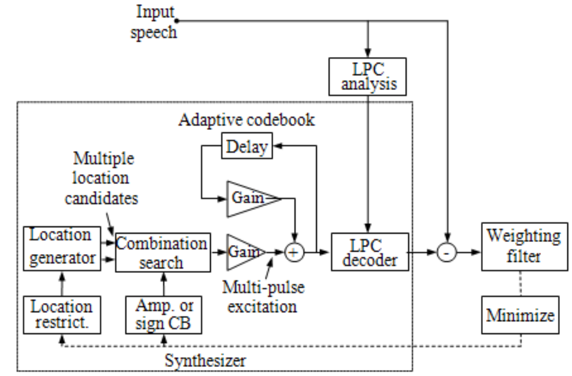

Fig. 2: MP-CELP core coder

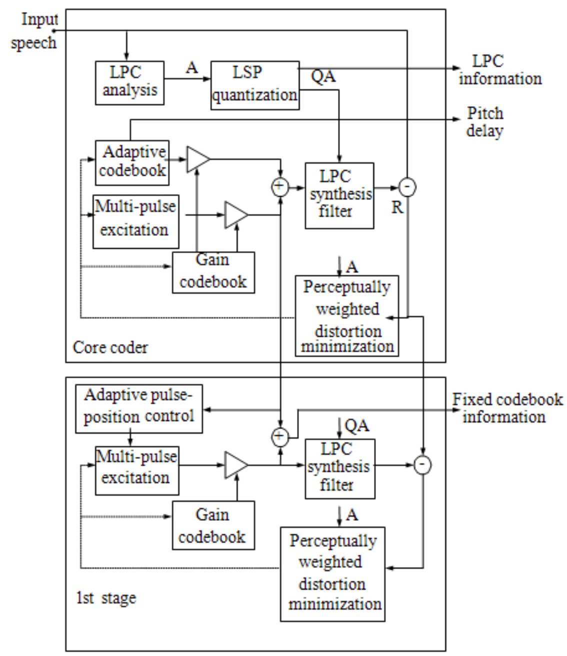

Fig. 3: 1-stage bitrate scalable MP-CELP coder

The pulse positions are restricted in the algebraic-structure codebook and determined by an analysis-by-synthesis approach (Laflamme *et al*., 1991). 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.

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 (Nomura *et al*., 1998). The bitrate scalable tool encodes the residual signal produced at the MP-CELP core coder utilizing the multi-pulse vector quantization.

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

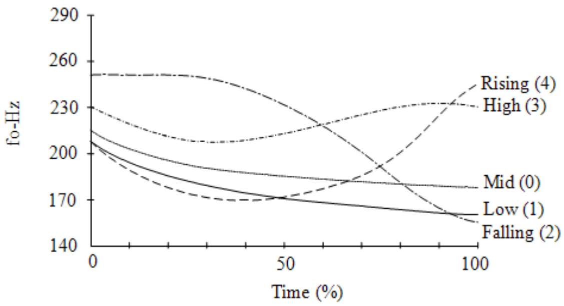

Fig. 4: Fundamental Frequency characteristic of 5 tones in Thai

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.

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. 4 (Chompun *et al*., 2001a; 2001b). 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.

Since pitch delay (inverse of f0) significantly involves in tone of tonal language, this study proposes an improvement of the bitrate scalable MP-CELP coder by applying the 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 (Chompun *et al*., 2001a; 2001b).

Gaussian Channel Model: In the AWGN channel, zero-mean white Gaussian noise is added to the transmitted signal s(t), so that the received signal r(t) can be represented as:

$$r(t) = s(t) + n(t) \tag{1}$$

where, n(t) is a zero-mean white Gaussian noise process with power N0/2 (Manglani and Bell, 2001).

Rayleigh fading channel model: Small scale fading is comprised of two independent mechanisms: the time spreading of the signal and the time varying behavior of the channel. A Doppler shift causes the time varying behavior of the channel. For cellular communications, a carrier frequency of fc = 1800 MHz and a vehicle speed of 45 mph results in a Doppler shift of fd = 120 Hz (Sklar, 1997).

RESULTS

Experimental conditions: The coding quality of the proposed coder was evaluated subjectively and types of simulated channels including AWGN channel and Rayleigh fading channel. The comparison tests objectively by using 36 tested sentences from 16 men and 16 women, some of them were shown in Table 2.

Table 2: Examples of Thai tested sentences (0, 1, 2, 3 and 4 at the end of each syllable represent Thai tones)

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

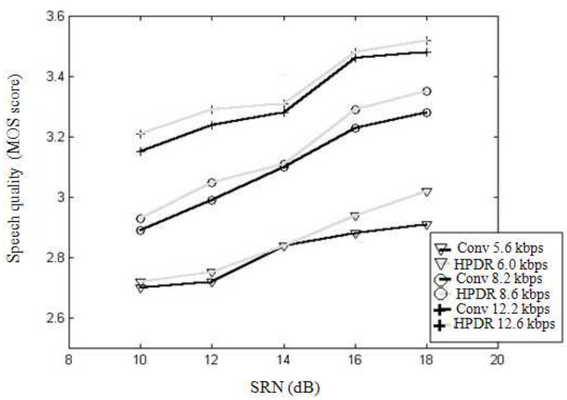

Fig. 5: Speech quality over AWGN channel

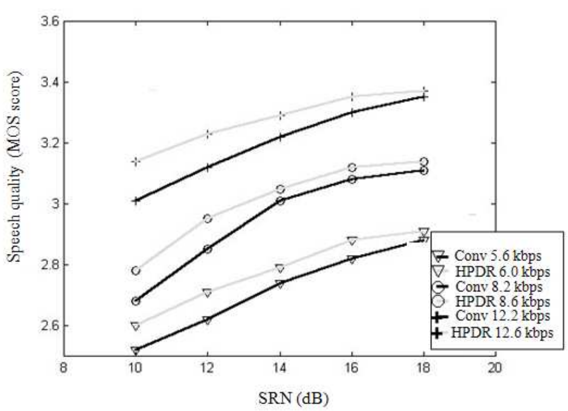

Fig. 6: Speech quality over Rayleigh fading channel

The effectiveness of the high pitch delay resolutions applied to the conventional coder was evaluated using average MOS scores. There are two between the conventional coder and the modified coder were conducted and shown in graphs of Fig. 5 and 6. Figure 5 show the speech quality transmitted through the AWGN channel, while Fig. 6 show the speech quality transmitted through the Rayleigh fading channel.

DISCUSSION

According to the subjective test (MOS score), graphs in Fig. 5 and 6 show that the speech quality of the coder modified with HPDR is above that of the conventional coder for all level of SNR at the same bitrate. This indicates that the proposed HPDR technique brings about better pitch precision which causes the improvement of the coding quality for tonal language over both AWGN channel and Rayleigh fading channel.

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 over both AWGN channel and Rayleigh fading channel.

REFERENCES

Adetunde, I.A. and B. Seidu, 2008. Dynamic response of loads on viscously damped axial force rayleigh beam. Am. J. Applied Sci., 5: 1110-1116. http://www.scipub.org/fulltext/ajas/ajas591110-1116.pdf

Chompun, S., S. Jitapunkul and D. Tancharoen, 2001a. Novel technique for tonal language speech compression based on a bitrate scalable MP-CELP coder. Proceeding of the IEEE International Conference on Information Technology: Coding and Computing, Apr. 2-4, IEEE Xplore Press, Las Vegas, NV., 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: 519-522. http://daisy.ee.eng.chula.ac.th/~d1oatty/oat_files/group_files/study/SCI2001_S_Chompun.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: 278-282. http://teacher.en.rmutt.ac.th/ktw/Resources/Full%20study%20PDF/SNLP%202000%20Proceedings.pdf

Chompun, S., 2004. Fine granularity scalability for MP-CELP based speech coding with HPDR technique. Proceeding of the IEEE Asia-Pacific Conference on Circuits and Systems, Dec. 6-9, IEEE Xplore Press, USA., pp: 197-200. DOI: 10.1109/APCCAS.2004.1412726

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

Manglani, M.J. and A.E. Bell, 2001. Wavelet modulation performance in Gaussian and Rayleigh fading channels. Proceeding of Military Communications Conference, Lockheed Martin Corporation and the Aerospace Corporation, Oct. 28-31, IEEE Xplore Press, Washington DC., USA., pp: 845-849. DOI: 10.1109/MILCOM.2001.985959

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, 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. 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

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, WA., USA., pp: 529-532. DOI: 10.1109/ICASSP.1998.674390

Sklar, B., 1997. Rayleigh fading channels in mobile digital communication system. I. Characterization. IEEE Commun. Mag., 35: 90-100. DOI: 10.1109/35.601747

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, Oct. 16-20, ISCA, Beijing, China, pp: 47-50. http://www.isca-speech.org/archive/icslp_2000/i00_3047.html

Vlasie, V. and M. Rousseau, 2005. Non destructive tests of structural bonds by guided ultrasonic waves: Effect of a surface pretreatment or a localized defect. Am. J. Applied Sci., 2: 739-745. DOI: 10.3844/ajassp.2005.739.745