American Journal of Applied Sciences 9 (3): 278-282, 2012
ISSN 1546-9239

Speech Compression for Noise-Corrupted Thai Dialects

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: Dialects of Thai are quite different in the speaking styles. Environmental noises play an important role in corrupting the communication speech quality. Two factors affect the coded speech in the present speech communication. It is necessary to investigate how the two factors influence on the speech compression. **Approach:** In this study, the Multi-Pulse based Code Excited Linear Predictive (MP-CELP) coder and the Conjugate Structure Algebraic Code Excited Linear Predictive (CS-ACELP) coder are selected as the coding methods. This study shows the effects of the six kinds of noise to speech coding quality. The comparison of speech quality of the four coded Thai dialects is conducted. The speech material includes a hundred male speech utterances and a hundred female speech utterances. Four speaking styles include Thai Northern, North Eastern, Southern and Central dialects. Five sentences of Thai speech are chosen. Six types of noise include train, factory, motorcycle, air conditioner, men speaker and women speaker. Moreover, five levels of each type of noise are varied from 0-20 dB. The subjective test of mean opinion score are exploited in the evaluation process. **Results:** The experimental results show that CS-ACELP gives better speech quality than that of MP-CELP at all three bitrates of 6000, 8600 and 12600 bps. When considering the levels of noise, the 20-dB noise gives the best speech quality, while 0-dB noise gives the worst speech quality. When considering the speech gender, male speech gives better results than that of female speech. When considering the types of dialect, the central dialect gives the best speech quality, while the North dialect gives the worst speech quality. Finally, when considering the types of noise, the air-conditioner noise gives the best speech quality, while the train noise gives the worst speech quality. **Conclusion:** From the study, it can be seen that coding method, type of noise, level of noise, speech gender and dialect influence on the coding speech quality.

Key words: Adaptive pulse, Conjugate Structure Algebraic Code Excited Linear Predictive (CS-ACELP), speech coding, bitrate scalability, Linear Prediction (LP), noise-corrupted speech, Thai dialects, coding rate, speech communication

INTRODUCTION

In Thai speech, there are four main dialects including Northern, North Eastern, Southern and Central dialects. They are mutually different in speaking style and prosody. Their prosodic information conveys the uniqueness of their intonations. Typically, their speaking rates are also varied, that is, the Northern dialect's rate is rather slow, while the Southern dialect's is the quickest rate among those four dialects.

In recent speech communication, low bitrate speech coding or compression is highly required to increase the channel capacity. Moreover, the flexibility in the coding rate is also necessary to support the surge of the traffic occupancies depending on the type and number of users. Speech compression is expected to manage these requirements (Chompun *et al*., 2000; Chomphan, 2010a; 2010b). Nowadays, the multimedia applications require special considerations for packet loss due to the existence of channel errors or noises in the noisy communication channel depicted in Fig. 1.

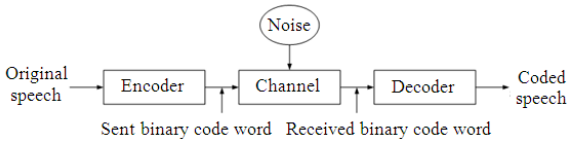

Fig. 1: Block diagram of noisy communication channel

To overcome this problem, a bitrate-scalable speech coder has been developed where the received speech signal can be decoded from the received packets, which contain only some of the whole encoded packets. The CS-ACELP algorithm was developed and standardized as ITU G.729 speech coder at the coding rate of 8 kbps in 1995. Thereafter, the MP-CELP speech coder has been developed to support the scalability functionality. The MP-CELP coder uses the multi-pulse excitation sequence which the number of pulses in fixed-entry codebook is selective for bitrate scalability functionality according to one of the MPEG-4 CELP speech coder requirements (Nomura *et al*., 1998; Chomphan, 2010b). In the MP-CELP speech encoder, amplitudes or signs (amplitude of +1 or -1) for generating the multi-pulse excitation are vector quantized. To improve speech quality for background noise environments, the adaptive pulse location restriction method are conducted (Ozawa and Serizawa, 1998). The coding rates are adjustable from 4-12 kbps by varying the number of pulses in the excitation sequence (Chomphan, 2010a).

This study proposes a study of the quality of speech coding based on the practical application which considers the communication environment with various types of background noises. Moreover we also considered the dialects of Thai speech with different speaking styles that may cause different speech quality with the same coding algorithm. Furthermore, the gender of speech and the levels of noise (in sense of signal to noise ratio) are also considered.

MATERIALS AND METHODS

CS-ACELP algorithm: The CS-ACELP coder is developed from the conventional Code-Excited Linear Predictive (CELP) coding algorithm (Chomphan, 2011a; 2011c). The coder extracts the speech features on the speech frames of 10 ms corresponding to 80 samples at a sampling rate of 8000 Hz. The extracted parameters of the CELP model include the linear-prediction filter coefficients, adaptive and fixed-codebook indices and gains. They are encoded and transmitted to the channel as shown in Fig 1. At the decoder, these parameters are employed to retrieve the excitation and synthesis filter parameters. The synthesized speech is reconstructed by filtering this excitation through the short-term synthesis filter based on a 10th order linear prediction filter and the long-term synthesis filter using adaptive-codebook approach. After obtaining the reconstructed speech, the speech is enhanced by filtering at a post processing unit.

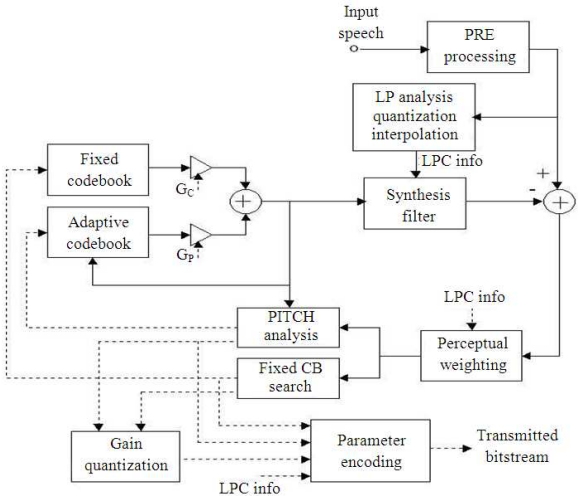

Fig. 2: Block diagram of CS-ACELP encoder

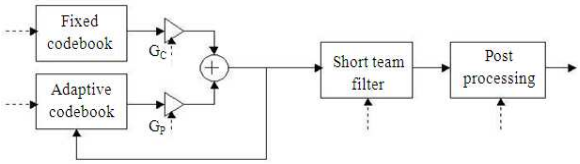

Fig. 3: Block diagram of CS-ACELP decoder

The block diagram of CS-ACELP encoder is shown in Fig. 2. The input signal is high-pass filtered and scaled in the pre-processing unit. Linear Prediction (LP) analysis is done every 10 ms frame to compute the LP coefficients. They are subsequently converted to Line Spectrum Pairs (LSP) and quantized using predictive two-stage vector quantization. The excitation is selected by applying an analysis-by-synthesis search procedure in which the error minimization between original and reconstructed speech is conducted.

The block diagram of CS-ACELP decoder is shown in Fig. 3. At the beginning, the parameters indices are extracted from the received bitstream. They are subsequently decoded to obtain the coder parameters for every 10-ms speech frame. The speech is reconstructed by filtering the excitation through the LP synthesis filter. Finally, the reconstructed speech signal is filtered at a post-processing unit which includes an adaptive post-filter, a high-pass filter and a scaling operation.

MP-CELP algorithm: The principle concepts for the bitrate scalable MP-CELP coder are explained in 2 parts of a core coder and a bitrate scalable tool (Chomphan, 2011a; 2011b; 2011c).

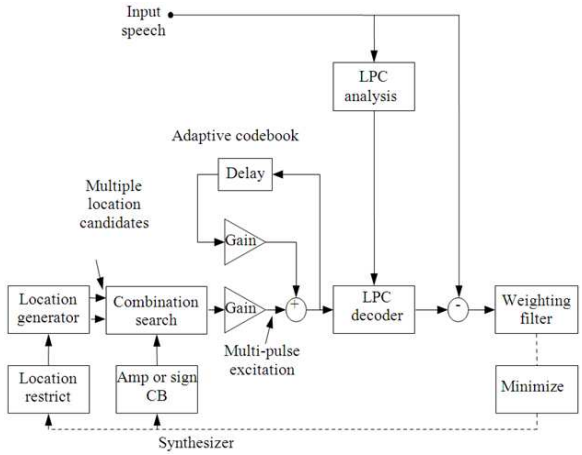

Fig. 4: Block diagram of MP-CELP core coder

A core coder: The core coder obtains the high coding performance by applying a multi-pulse vector quantization as depicted in Fig. 4 (Taumi *et al*., 1996; Ozawa *et al*., 1997). The input speech of a 10-ms frame is analyzed at the LP and pitch analysis. The LP coefficients are then quantized in the Line Spectrum Pairs (LSP) domain. The pitch delay is simultaneously 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 is composed of several non-zero pulses. Their pulse positions are restricted in the algebraic-structure codebook and calculated by an analysis-by-synthesis scheme, e.g., (Laflamme *et al*., 1991). The pulse signs and positions are subsequently encoded, while the gains for pitch predictor and the multi-pulse excitation are normalized by the frame energy and also encoded.

A bitrate scalable tool: Three stages of the bitrate scalable tools are conducted. It is embedded adjacently to the core coder as depicted in Fig. 5. The tool encodes the residual signal from the core coder utilizing the multi-pulse vector quantization. An adaptive pulse position control is applied 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 core coder or the previous stage. The pulse positions are selected 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 similar to that of the core coder (Chomphan, 2011a; 2011b; 2011c).

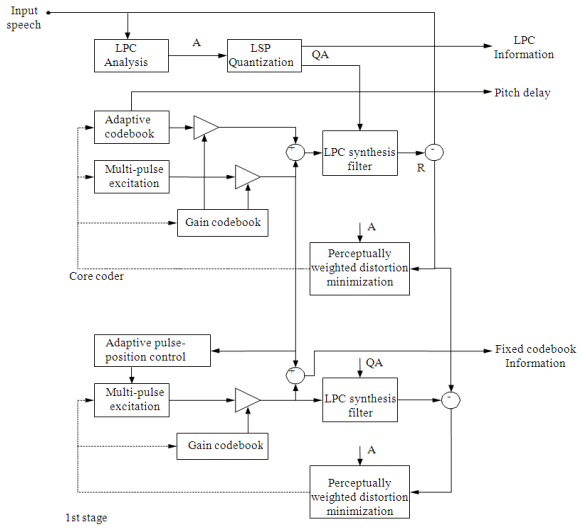

Fig. 5: Block diagram of one-stage bitrate scalable MP-CELP coder

RESULTS

In this study, the evaluation results are concentrated on speech coding for noise-corrupted Thai dialects by using two coding methods of CS-ACELP and MP-CELP. The selected bitrates of MP-CELP are 6000, 8600 and 12600 bps. The speech material includes a hundred of male speech utterances and a hundred of female speech utterances. Four speaking styles include Thai Northern, North Eastern, Southern and Central dialects. Five sentences of Thai speech are chosen. Six types of noise include train, factory, motorcycle, air conditioner, men speaker and women speaker. As for level of noise, five levels of each type of noise are varied from 0-20 dB. The subjective test of mean opinion score are exploited in the evaluation process.

The results are summarized in the following Fig. 6-10.

DISCUSSION

From the experimental results, it can be concluded that CS-ACELP gives the better speech quality than that of MP-CELP at all three bitrates of 6000, 8600 and 12600 bps as seen in Fig. 6 and 7. Moreover, when considering the different dialects in Fig. 6, the Central dialect gives the best speech quality, while the North dialect gives the worst speech quality. When considering the types of noises, the air conditioner noise gives the best speech quality, while the train noise gives the worst speech quality as seen in Fig. 8.

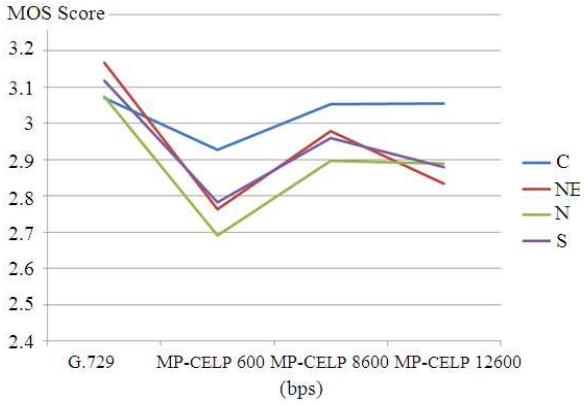

Fig. 6: MOS scores for different coding methods with four Thai dialects. (C, NE, N and S denote Central, North East, North and South dialects, respectively)

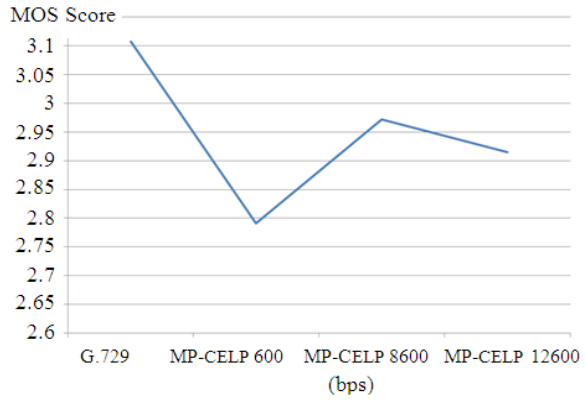

Fig. 7: MOS scores for different coding methods

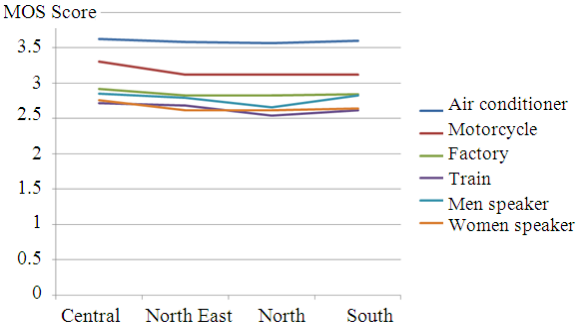

Fig. 8: MOS scores for all four Thai dialects with six different types of noises

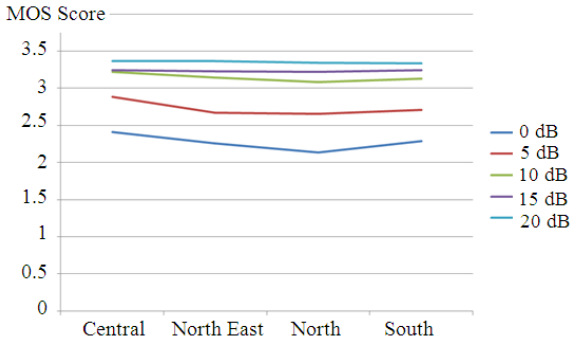

Fig. 9: MOS scores for all four Thai dialects with five different levels of noises

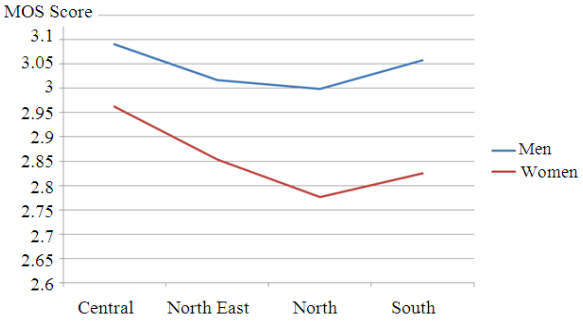

Fig. 10: MOS scores for all four Thai dialects with two genders

When considering the levels of noises, the 20-dB noise gives the best speech quality, while 0-dB noise gives the worst speech quality as seen in Fig. 9. Finally, when considering the speech gender, male speech gives the better results than that of female speech as seen in Fig. 10.

CONCLUSION

This study proposes a study of speech compression for noise-corrupted Thai dialects by using two coding methods of CS-ACELP and MP-CELP. The experimental results show that CS-ACELP gives the better speech quality than that of MP-CELP at all three bitrates. When considering the different dialects, the Central dialect mostly gives the best speech quality, while the North dialect mostly gives the worst speech quality. When considering the levels of noises, the 20-dB noise gives the best speech quality, while 0-dB noise gives the worst speech quality. When considering the speech gender, male speech gives the better results than that of female speech. Finally, when considering the types of noises, the air

conditioner noise gives the best speech quality, while the train noise gives the worst speech quality.

ACKNOWLEDGEMENT

The reaearchers is grateful to Kasetsart University at Si Racha campus for the research scholarship through the board of research.

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.2010.1438.1442

Chomphan, S., 2011a. Analysis of fundamental frequency contour of coded speech based on multi-pulse based code excited linear prediction algorithm. J. Comput. Sci., 7: 865-870. DOI: 10.3844/jcssp.2011.865.870

Chomphan, S., 2011c. Tonal language speech compression based on a bitrate scalable multi-pulse based code excited linear prediction coder. J. Comput. Sci., 7: 154-158. DOI: 10.3844/jcssp.2011.154.158

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, (NLP' 00), Chiangmai, Thailand, pp: 1-5.

Chomphan, S., 2011b. Speech compression for noise-corrupted thai expressive speech. J. Comput. Sci., 7: 1565-1573. DOI: 10.3844/jcssp.2011.1565.1573

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, May 12-15, IEEE Xplore Press, Seattle, 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: Fundamental 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. Proceedings 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

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, USA., pp: 562-565. DOI: 10.1109/ICASSP.1996.541158