Journal of Computer Science 7 (10): 1565-1573, 2011
ISSN 1549-3636
© 2011 Science Publications

Speech Compression for Noise-Corrupted Thai Expressive Speech

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 speech communication, speech coding aims at preserving the speech quality with lower coding bitrate. When considering the communication environment, various types of noises deteriorates the speech quality. The expressive speech with different speaking styles may cause different speech quality with the same coding method. **Approach:** This research proposed a study of speech compression for noise-corrupted Thai expressive speech by using two coding methods of CS-ACELP and MP-CELP. The speech material included a hundredmale speech utterances and a hundred female speech utterances. Four speaking styles included enjoyable, sad, angry and reading styles. Five sentences of Thai speech were chosen. Three types of noises were included (train, car and air conditioner). Five levels of each type of noise were varied from 0-20 dB. The subjective test of mean opinion score was exploited in the evaluation process. **Results:** The experimental results showed that CS-ACELP gave the better speech quality than that of MP-CELP at all three bitrates of 6000, 8600-12600 bps. When considering the levels of noise, the 20-dB noise gave the best speech quality, while 0-dB noise gave the worst speech quality. When considering the speech gender, female speech gave the better results than that of male speech. When considering the types of noise, the air-conditioner noise gave the best speech quality, while the train noise gave the worst speech quality. **Conclusion:** From the study, it can be seen that coding methods, types of noise, levels of noise, speech gender influence on the coding speech quality.

Key words: Multi-Pulse based (MP-CELP), code excited, linear predictive, speech coding, bitrate scalability, Linear Prediction (LP), expressive speech, speech compression, speech quality, coding methods, speech signal, perceptual weighting

INTRODUCTION

In recent speech communication network, low bitrate speech compression is highly required to preserve the channel capacity. The flexibility of coding rate are also needed to support the variety of the traffic occupancies depending on the type and number of users. Signal compression or speech coding aims to perform this (Chompun *et al*., 2000; Chomphan, 2010a; 2010b). Presently, the multimedia applications such as videophone and teleconferencing on ATM and Internet are considerably interested, the high quality speech coders with low bitrates are highly demanded (Chompun *et al*., 2000). These applications require special considerations for packet loss. To relief this problem, a bitrate-scalable speech coder has been studied where the synthesized speech signal can be decoded from the received packets, which contain only some of the whole encoded packets. In 1995, Conjugate-Structure Algebraic Code Excited Linear Predictive (CS-ACELP) coding was developed and standardized as ITU G.729 speech coding at the coding rate of 8 kbps. A few years later, MP-CELP speech coder has been developed to be a scalable coder. With the flexible functionality, this 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 (Ozawa *et al*., 1996; Chomphan, 2010b) In the MP-CELP speech coder, amplitudes or signs for generating the multi-pulse excitation are vector quantized simultaneously. Moreover, to improve speech quality for background noise conditions, the adaptive pulse location restriction method are applied (Ozawa and Serizawa, 1998). The speech coder operates at various bitrates ranging from 4-12 kbps utilizing the flexibility in multi-pulse excitation coding (Chomphan, 2010a).

This study proposes a study of the quality of speech compression based on the practical usage which considers the communication environment with various types of noises. Moreover we also considered the expressive speech with different speaking styles that may cause different speech quality with the same coding method. Furthermore, the gender of speech and the levels of noise (in sense of signal to noise ratio) are also studied (Nadia *et al*., 2009; Tan *et al*., 2009).

MATERIALS AND METHODS

CS-ACELP algorithm: The CS-ACELP coder is based on the Code-Excited Linear Predictive (CELP) coding model. The coder operates on speech frames of 10 ms corresponding to 80 samples at a sampling rate of 8000 samples per- sec. For every 10 m sec frame, the speech signal is analyzed to extract the parameters of the CELP model (linear-prediction filter coefficients, adaptive and fixed-codebook indices and gains). These parameters are encoded and transmitted. At the decoder, these parameters are used to retrieve the excitation and synthesis filter parameters. The speech is reconstructed by filtering this excitation through the short-term synthesis filter based on a 10th order liner prediction filter and the long-term or pitch synthesis filter implemented using adaptive-codebook approach. After computing the reconstructed speech, it is further enhanced by a post-filter.

The encoding principle is shown in Fig. 1. The input signal is high-pass filtered and scaled in the pre-processing block. The pre-processing signal serves as the input signal for all subsequent analysis. LP analysis is done once per 10 ms frame to compute the LP coefficients. These coefficients are converted to Line Spectrum Pairs (LSP) and quantized using predictive two-stage vector quantization with 18 bits. The excitation signal is chosen by using an analysis-by-synthesis search procedure in which the error between original and reconstructed speech is minimized according to a perceptually weighted distortion measure. This is done by filtering the error signal with a perceptual weighting filter, whose coefficients are derived from the unquantized LP filter. The amount of perceptual weighting is made adaptive to improve the performance for input signals with a flat frequency-response. The decoder principle is shown in Fig. 2.

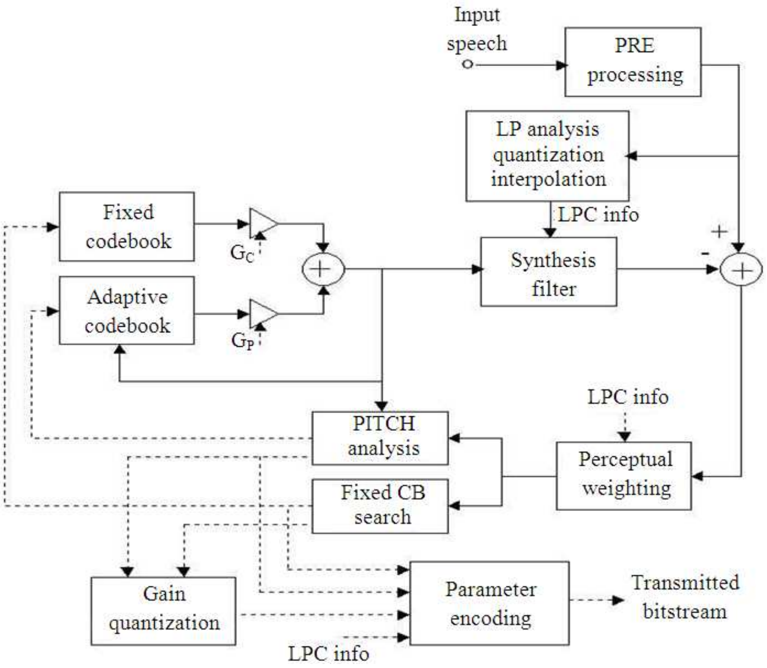

Fig. 1: Block diagram of CS-ACELP encoder

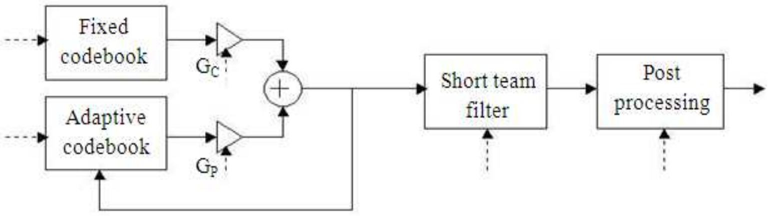

Fig. 2: Block diagram of CS-ACELP decoder

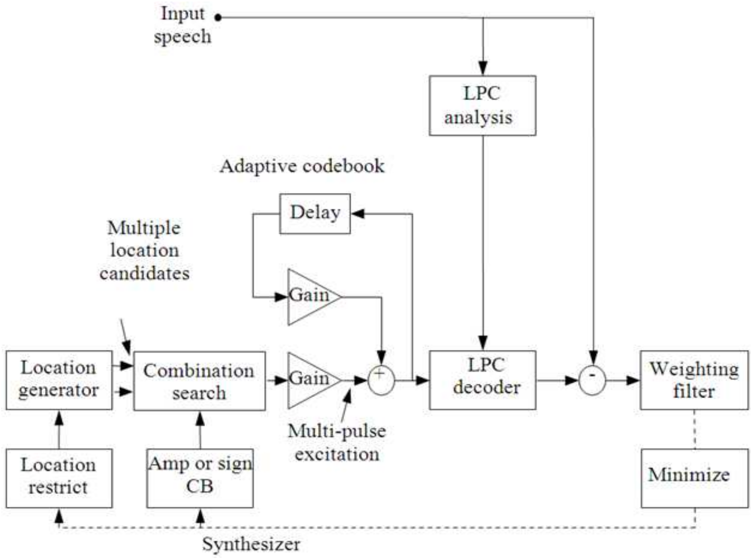

Fig. 3: Block diagram of MP-CELP core coder

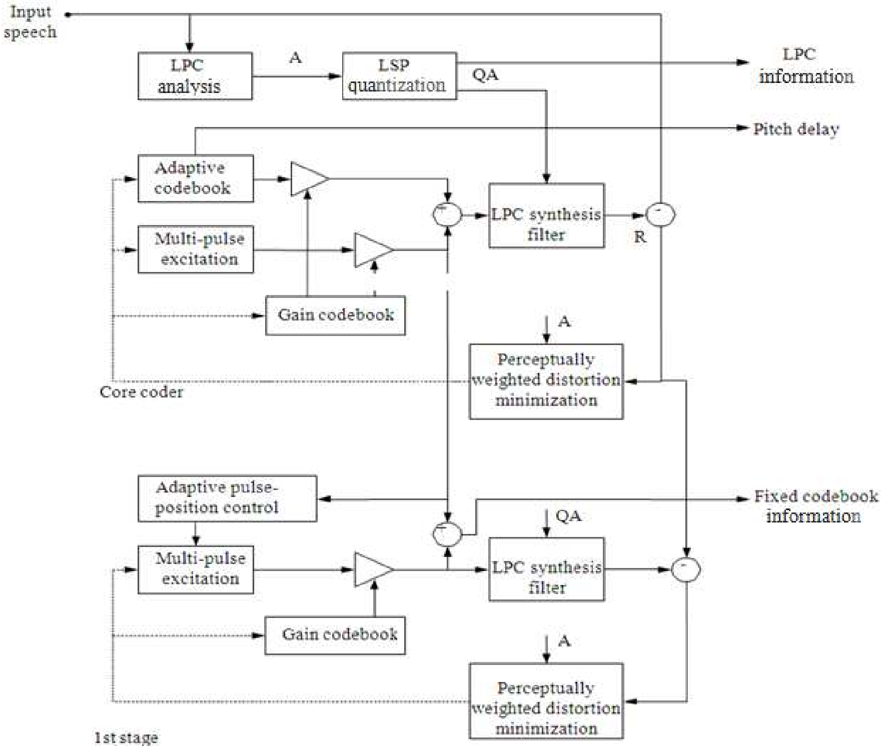

Fig. 4: Block diagram of one-stage bitrate scalable MP-CELP coder

First, the parameters indices are extracted from the received bitstream. These indices are decoded to obtain the coder parameters corresponding to a 10 ms speech frame. These parameters are the LSP coefficients, the 2 fractional pitch delays, the 2 fixed-codebook vectors and the 2 sets of adaptive and fixed-codebook gains. The LSP coefficients are interpolated and converted to LP coefficients for each subframe. Then, for each 5 ms subframe, the excitation is constructed by adding the adaptive and fixed-codebook vectors scaled by their respective gains, the speech is reconstructed by filtering the excitation through the LP synthesis filter, finally, the reconstructed speech signal is passed through a post-processing stage, which includes an adaptive post-filter based on the long-term and short-term synthesis filter, followed by a high-pass filter and scaling operation.

Voice Activity Detection is in the pre-processing part to decide the input speech frame as voiced or unvoiced speech. Consequently, the unvoiced speech mode neglects the adaptive codebook quantization part because no periodicity is needed while the voiced speech mode still employs both fixed and adaptive quantization part.

MP-CELP algorithm: The operation principle for bitrate scalable MP-CELP coder can be divided into 2 parts, the MP-CELP core coder and the 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. 3 (Ozawa *et al*., 1996). The input speech of a 10-ms-length 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). 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, subsequently.

Bitrate scalable tool: This study applies 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. 4. 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 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, 5-10. The bit allocation is shown in Table 1. As for bitrate scalable tool, each stage increases the bitrate of 800 bps. Though, as for one multi-pulse, the total bitrate are 5600, 6400, 7200-8000 bps respectively. As for five multi-pulses, the total bitrate are 8200, 9000, 9800-10600 bps respectively. And as for ten multi-pulses, the total bitrate are 12200, 13000, 13800-14600 bps respectively.

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

RESULTS

In the evaluation results mainly focus on speech compression for noise-Corrupted Thai expressive speech by using two coding methods of CS-ACELP and MP-CELP. The MP-CELP with three levels of bitrate scalability is selected as the core speech coder. The selected bitrates are 5600, 8200-12200 bps. The speech material includes a hundred of male speech utterances and a hundred of female speech utterances. Four speaking styles include enjoyable, sad, angry and reading styles. Five sentences of Thai speech are chosen. Three types of noise include train, car and air conditioner. 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. The results are summarized in the following Fig. 5-29.

DISCUSSION

The experimental results show that CS-ACELP gives the better speech quality than that of MP-CELP at all three bitrates of 6000, 8600-12600 bps as seen in most of Fig. 5-28.

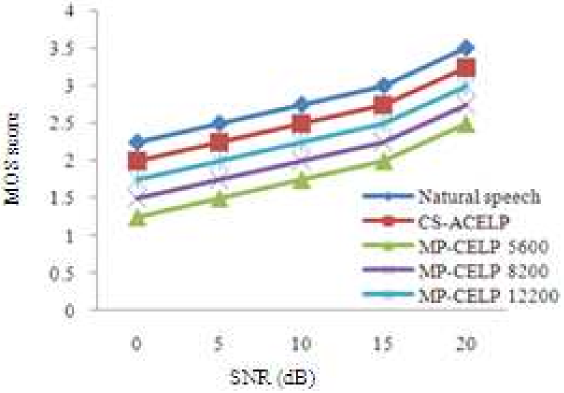

Fig. 5: MOS score of male angry speech with air-conditioner noise

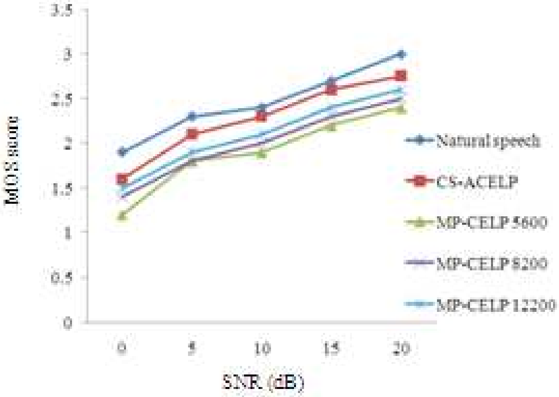

Fig. 6: MOS score of male angry speech with car noise

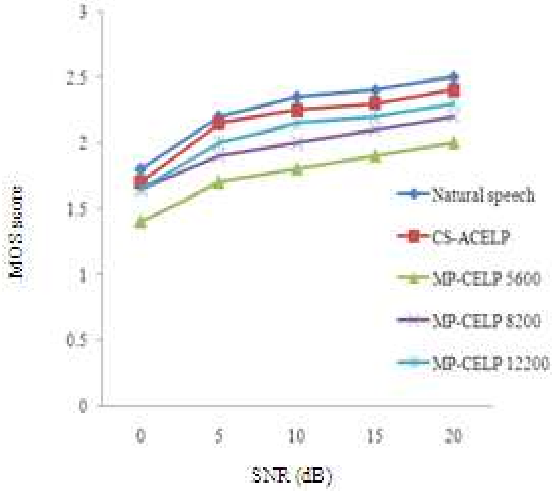

Fig. 7: MOS score of male angry speech with train noise

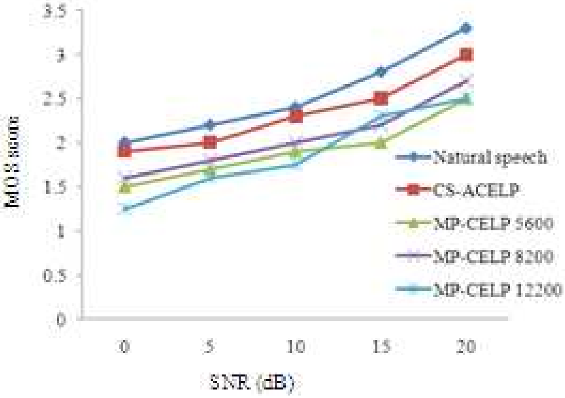

Fig. 8: MOS score of male enjoyable speech with air-conditioner noise

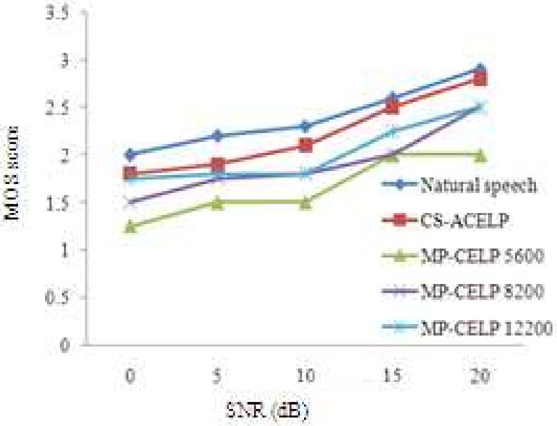

Fig. 9: MOS score of male enjoyable speech with car noise

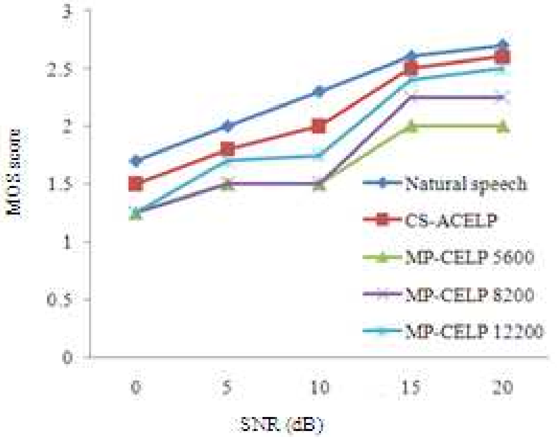

Fig. 10: MOS score of male enjoyable speech with train noise

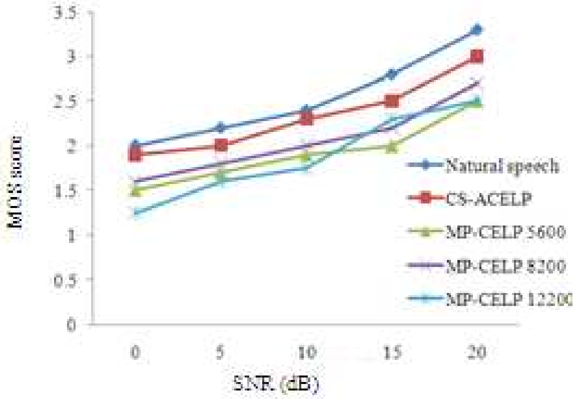

Fig. 11: MOS score of male reading speech with air-conditioner noise

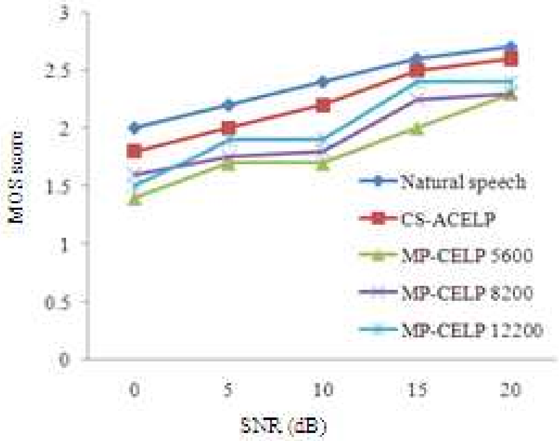

Fig. 12: MOS score of male reading speech with car noise

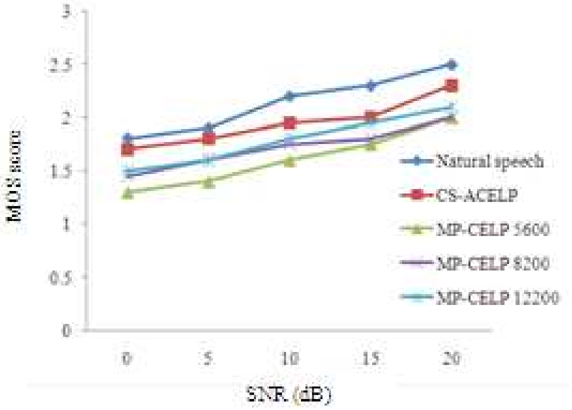

Fig. 13: MOS score of male reading speech with train noise

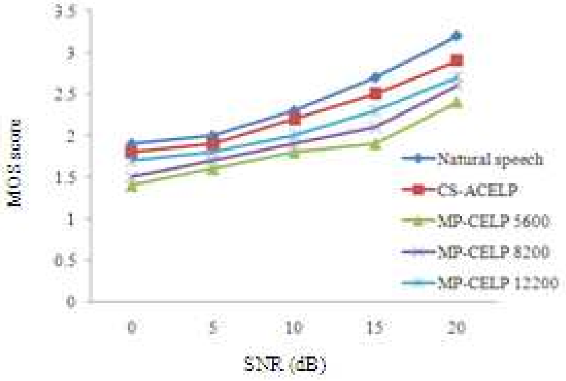

Fig. 14: MOS score of male sad speech with air-conditioner noise

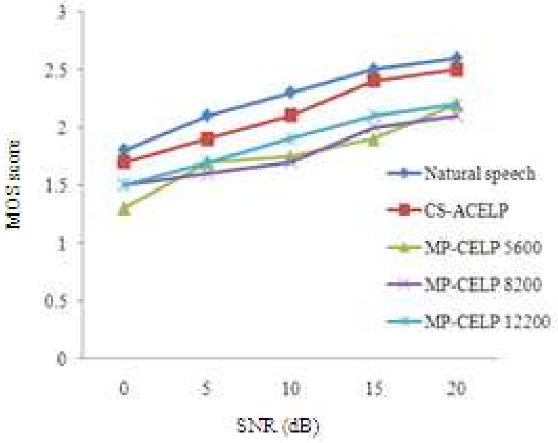

Fig. 15: MOS score of male sad speech with car noise

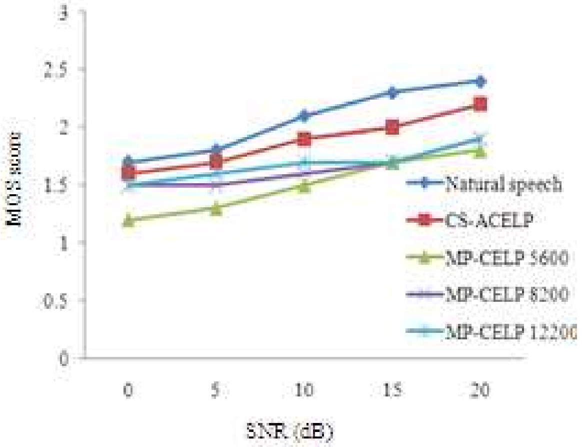

Fig. 16: MOS score of male sad speech with train noise

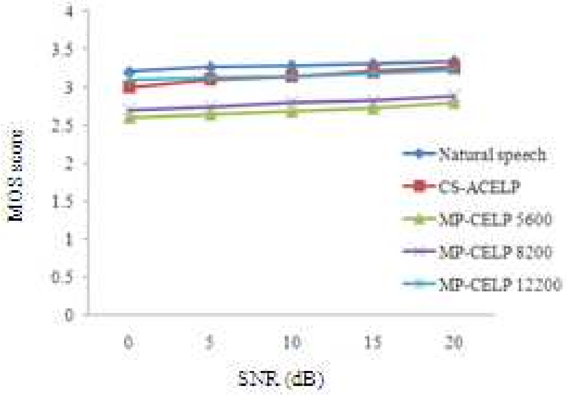

Fig. 17: MOS score of female angry speech with air-conditioner noise

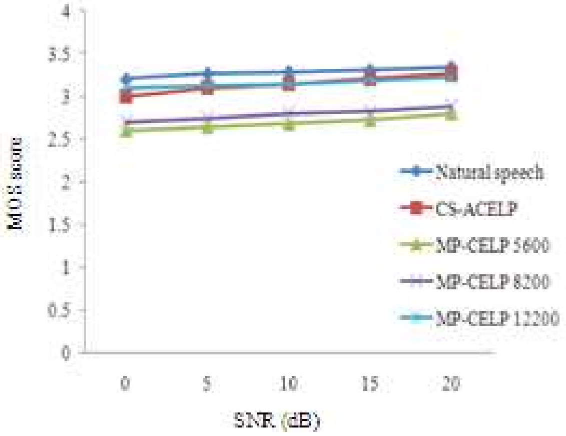

Fig. 18: MOS score of female angry speech with car noise

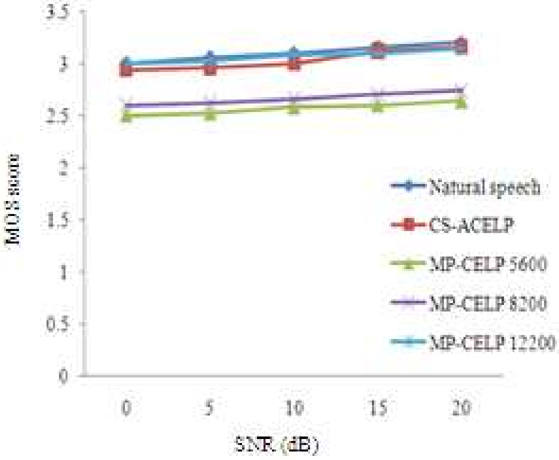

Fig. 19: MOS score of female angry speech with train noise

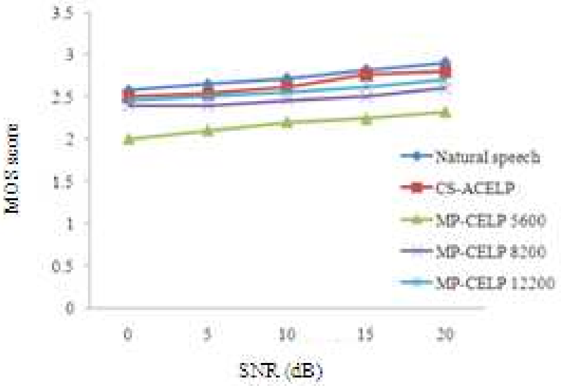

Fig. 20: MOS score of female enjoyable speech with air-conditioner noise

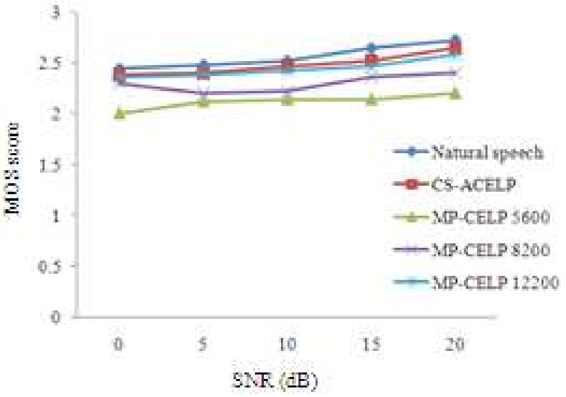

Fig. 21: MOS score of female enjoyable speech with car noise

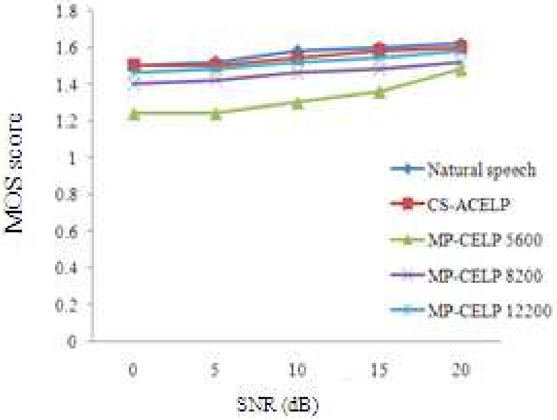

Fig. 22: MOS score of female enjoyable speech with train noise

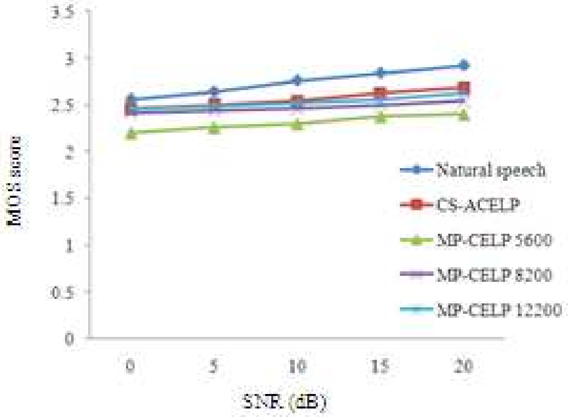

Fig. 23: MOS score of female reading speech with air-conditioner noise

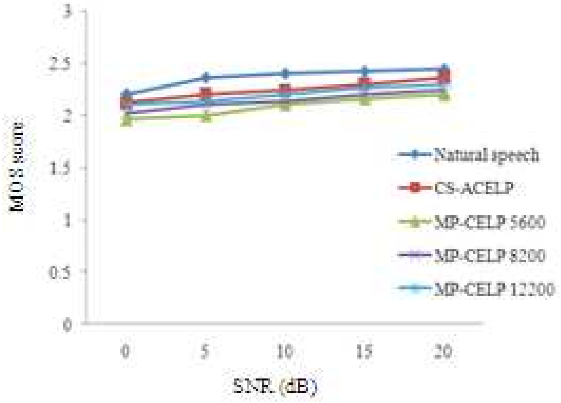

Fig. 24: MOS score of female reading speech with car noise

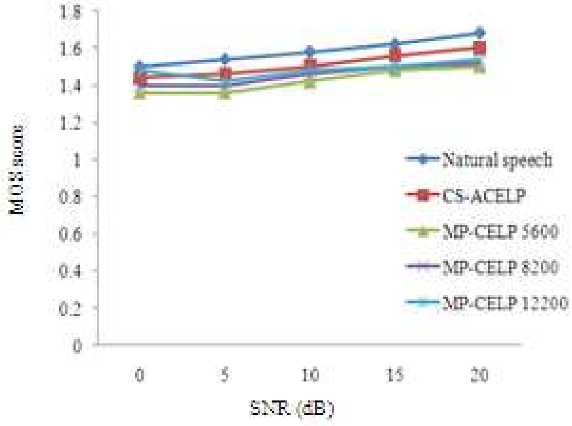

Fig. 25: MOS score of female reading speech with train noise

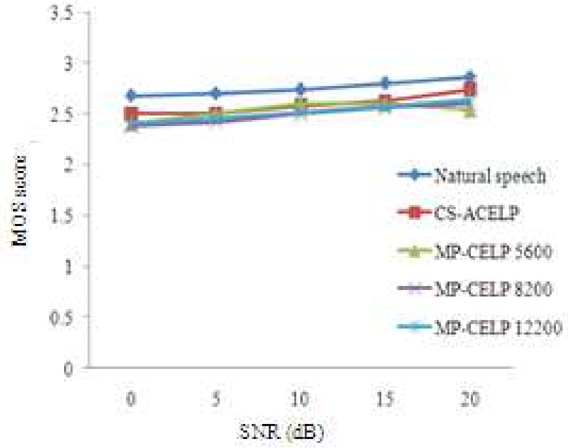

Fig. 26: MOS score of female sad speech with air-conditioner noise

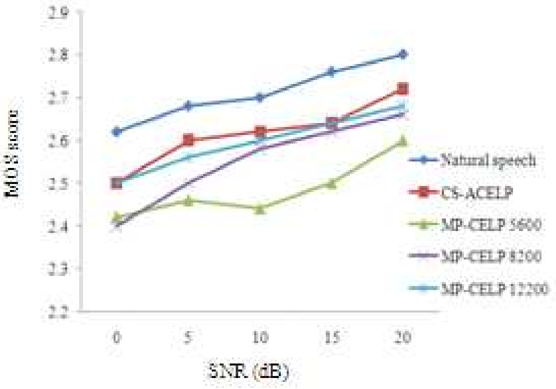

Fig. 27: MOS score of female sad speech with car noise

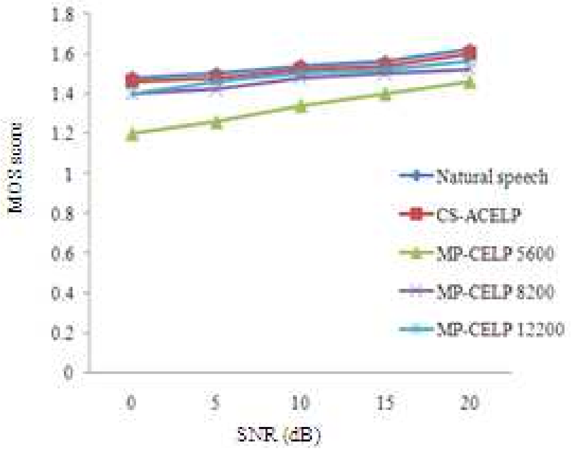

Fig. 28: MOS score of female sad speech with train noise

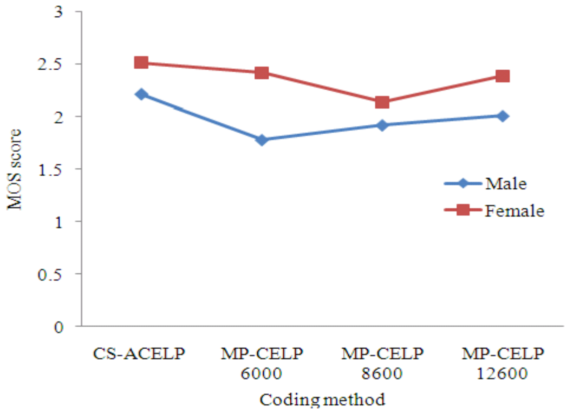

Fig. 29: MOS score of averaged male and female speech

When considering the levels of noise, the 20-dB noise gives the best speech quality, while 0-dB noise gives the worst speech quality as seen in most of Fig. 5-28. When considering the speech gender, female speech gives the better results than that of male speech as seen in most of Fig. 29. 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 as seen in most of Fig. 5-28.

CONCLUSION

This study proposes a study of speech compression for noise-Corrupted Thai expressive speech 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 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, female gives the better results than that of male speech. 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.

ACKNOWLEDGEMENT

The researchers are 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.2010.1433.1437

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'2000), Chiangmai, Thailand, pp: 1-6.

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 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

Al-Saidi, N.M.G. and M.R.M. Said, 2009. A new approach in cryptographic systems using fractal image coding. J. Math. Stat., 5: 183-189. DOI: 10.3844/JMSSP.2009.183.189

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. 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, USA, pp: 529-532. DOI: 10.1109/ICASSP.1998.674484

Ozawa, K., T. Nomura and M. Serizawa, 1997. MP-CELP speech coding based on multi-pulse vector quantization and fast search. IEICE Trans., 80: 55-63. DOI: 10.1002/(SICI)1520-6440(199711)80:11<55::AID-ECJC6>3.0.CO;2-R

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

Tan, T.S. and S. Hussain, 2009. Corpus design for malay corpus-based speech synthesis system. Am. J. Applied Sci., 6: 696-702. DOI: 10.3844/AJASSP.2009.696.702