American Journal of Applied Sciences 9 (8): 1321-1328, 2012
ISSN 1546-9239

Speech Compression of Thai Dialects with Low-Bit-Rate Speech Coders

[1, 2]Suphattharachai Chomphan
[1]Department of Electrical Engineering, Faculty of Engineering at Si Racha, Kasetsart University, 199 M.6, Tungsukhla, Si Racha, Chonburi, 20230, Thailand
[2]Center for Advanced Studies in Industrial Technology, Kasetsart University, 50 Ngam Wong Wan Rd, Ladyaow, Chatuchak, Bangkok, 10900, Thailand

Abstract: Problem statement: In modern speech communication at low bit rate, speech coding deteriorates the characteristics of the coded speech significantly. Considering the dialects in Thai, the coding quality of four main dialects spoken by Thai people residing in four core region including central, north, northeast and south regions has not been studied. **Approach:** This study presents a comparative study of the coding quality of four main Thai dialects by using different low-bit-rate speech coders including the Conjugate Structure Algebraic Code Excited Linear Predictive (CS-ACELP) coder and the Multi-Pulse based Code Excited Linear Predictive (MP-CELP) coder. The objective and subjective tests have been conducted to evaluate the coding quality of four main dialects. **Results:** From the experimental results, both tests show that the coding quality of North dialect is highest, meanwhile the coding quality of Northeast dialect is lowest. Moreover, the coding quality of male speech is mostly higher than that of female speech. **Conclusion:** From the study, it can be obviously seen that the coding quality of all Thai dialects are different.

Key words: Thai dialects, Multi-Pulse based Code Excited Linear Predictive (MP-CELP), Conjugate Structure Algebraic Code Excited Linear Predictive (CS-ACELP), objective test, subjective

INTRODUCTION

In the present day, the digital communication have been are considerably improved and developed. The audio, still image, video or text information can be transmitted through wire and wireless networks, meanwhile, the number of users to access these networks increases extremely. Therefore, the channel capacity must be increased, signal compression aims at overcoming this situation (Chompun *et al*., 2000).

In the communication system with an occurring of packet loss, the high quality speech compression or speech coder is highly demanded. One of standardization activities is conducted under the project of MPEG-4 (Nomura *et al*., 1998; Chomphan, 2010b). The MP-CELP coder has been proposed to be a scalable coder. This speech coder employs the multi-pulse excitation which the number of pulses in fixed-entry codebook is selective for bitrate scalability and multiple bitrates functionality according to the MPEG-4 CELP speech coder requirements, (Nomura *et al*., 1998; Chomphan, 2010b). It should be noted that the MP-CELP coder has been developed from the CS-ACELP coder standardized as 8-kb/s G.729 in 1995.

In the MP-CELP core coder, amplitudes or signs for multi-pulse excitation are simultaneously vector quantized. Moreover, to improve speech quality for background noise conditions, the adaptive pulse location restriction method are utilized (Ozawa and Serizawa, 1998). This coder operates at various bitrates ranging from 4-12 kbps by applying the flexibility in multi-pulse excitation coding (Chomphan, 2010a).

This study performs a comparative study of the coding quality of four main Thai dialects of spoken by Thai people residing in four core region including central, north, northeast and south regions. The core compression is based on two different low-bit-rate speech coders including the Conjugate Structure Algebraic Code Excited Linear Predictive (CS-ACELP) coder and the Multi-Pulse based Code Excited Linear Predictive (MP-CELP) coder. The objective and subjective tests have been performed to evaluate the coding quality of four main dialects.

MATERIALS AND METHODS

CS-ACELP coder: The CS-ACELP coder is based on the Code-Excited Linear Predictive (CELP) coding model.

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

The coder operates on speech frames of 10 ms corresponding to 80 samples at a sampling rate of 8000 samples per second. For every 10 ms 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 (Schroder and Sherif, 1997).

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. First, the parameters indices are extracted from the received bitstream.

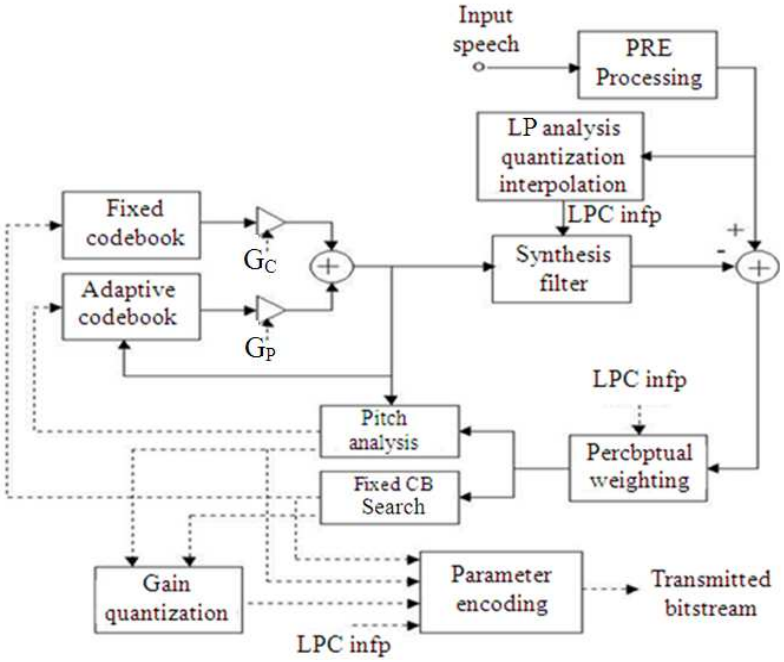

Fig. 1: Block diagram of CS-ACELP (G.729) encoder

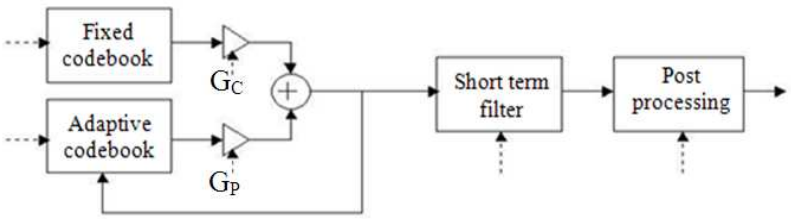

Fig. 2: Block diagram of CS-ACELP (G.729) decoder

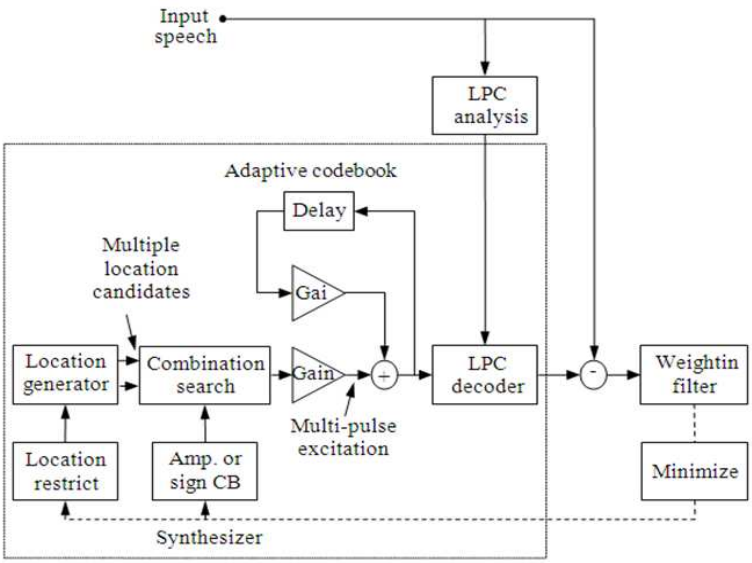

Fig. 3: Block diagram of MP-CELP coder

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.

MP-CELP coder: The MP-CELP core coder achieves a high coding performance by introducing a multi-pulse vector quantization as depicted in Fig. 3 (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. The pulse positions are restricted in the algebraic-structure codebook and determined by an analysis-by-synthesis approach, (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. The supporting core bitrates of this coder are 5600, 8200, 12200 bps, for the core coder with one pulse in fixed codebook, five pulses in fixed codebook and ten pulses in fixed codebook, respectively.

RESULTS

The coding quality of the bitrate-scalable coder was evaluated subjectively and objectively by using 50 tested sentences for each of four main dialects spoken by Thai people residing in four core region including central, north, northeast and south regions. Each dialect consists of the speech from a man and a woman.

As for the subjective test, The Mean Opinion Score (MOS) has been chosen for evaluating the coding quality of all dialects and genders. The subjects consist of four men and four women. The averaged MOS scores for each dialect are presented in Fig. 4-7.

As for the objective test, the Signal to Noise Ratio (SNR) score has been chosen for evaluating the coding quality of all dialects and genders to confirm the result from the subjective test. The SNR score are computed from the energy of natural speech and the energy of the difference between the natural speech and the encoded speech. The averaged SNR scores for each dialect are presented in Fig. 8-11.

Finally, Fig. 12-13 present the comparisons between the coding quality among all four dialects using the subjective test and the objective test, respectively.

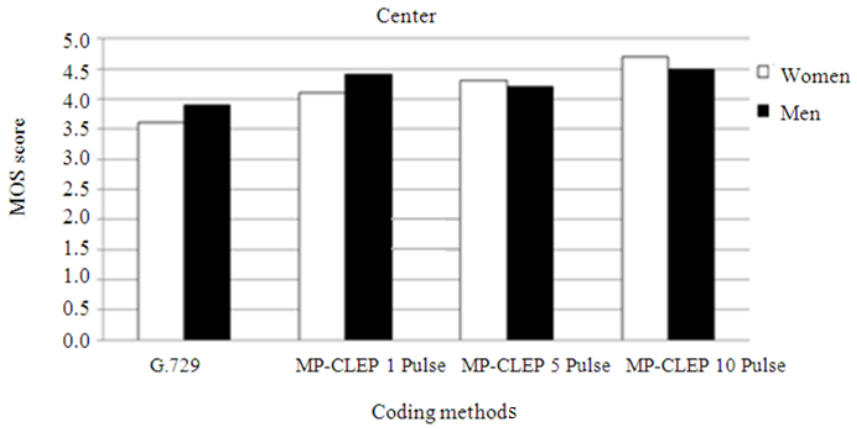

Fig. 4: Averaged MOS scores of different coding methods for center dialect

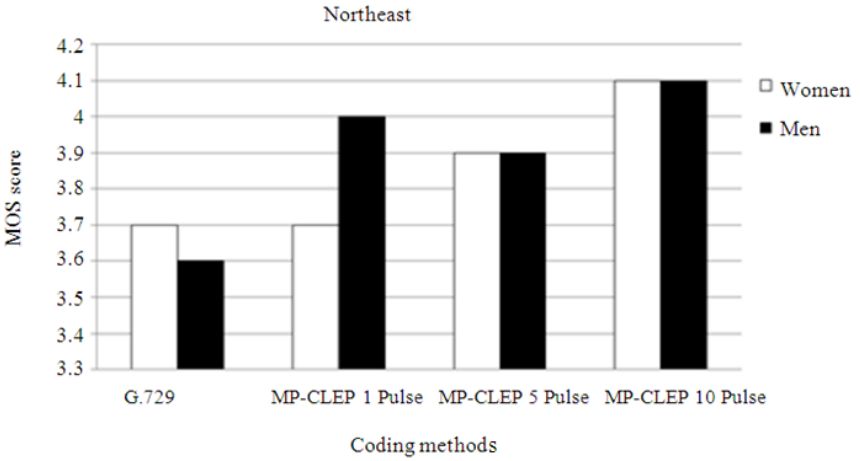

Fig. 5: Averaged MOS scores of different coding methods for Northeast dialect

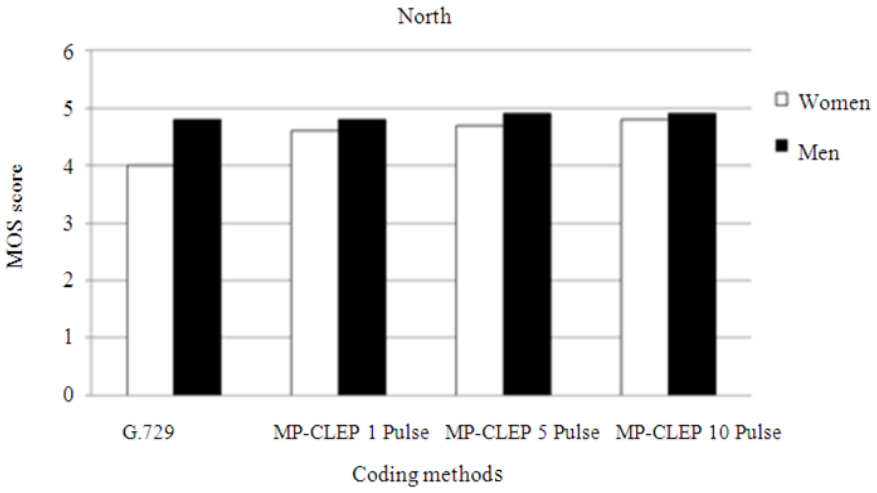

Fig. 6: Averaged MOS scores of different coding methods for North dialect

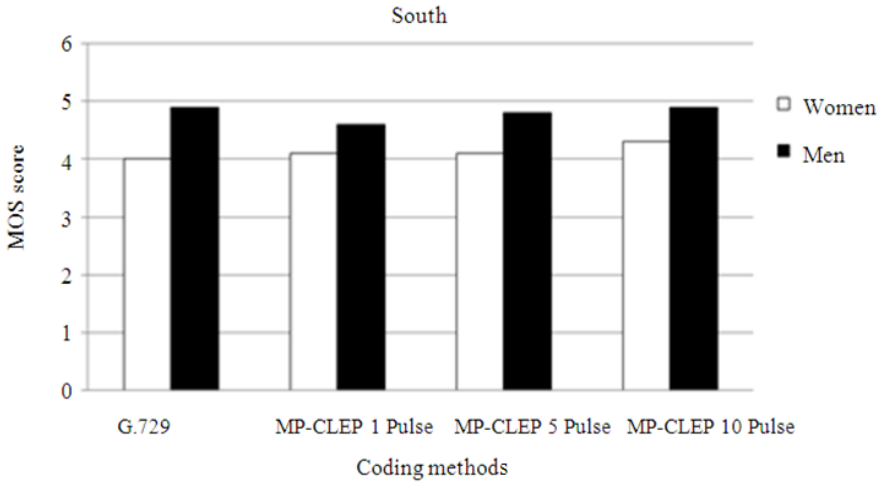

Fig. 7: Averaged MOS scores of different coding methods for South dialect

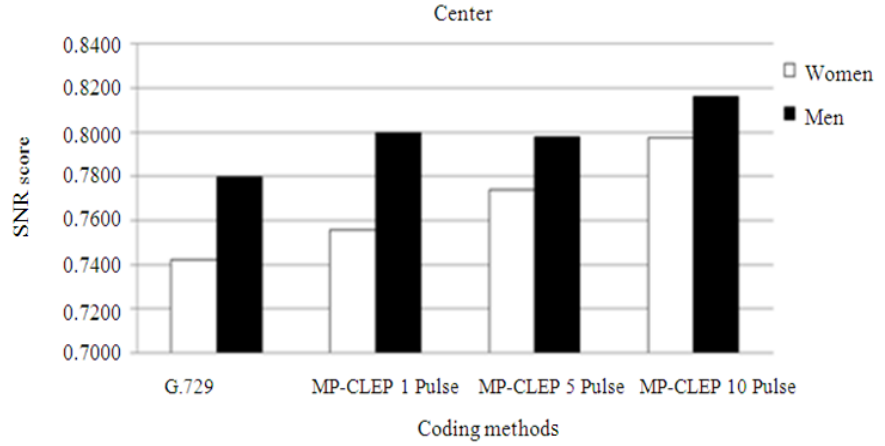

Fig. 8: Averaged SNR scores of different coding methods for Center dialect

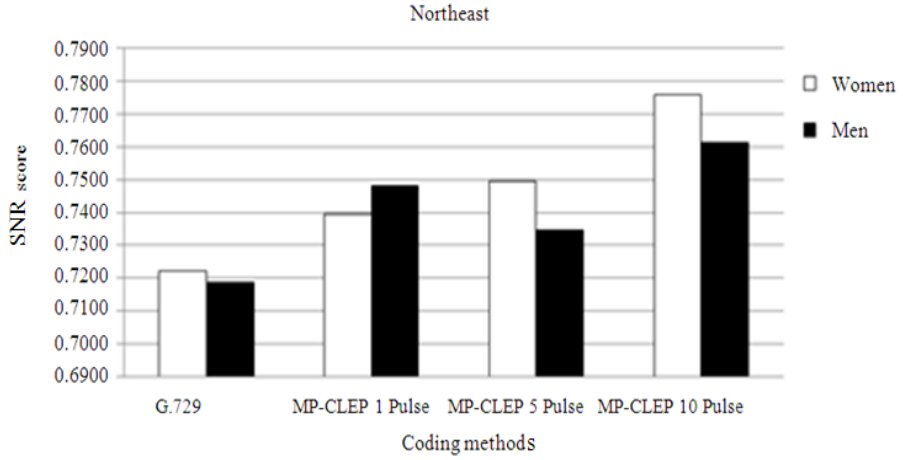

Fig. 9: Averaged SNR scores of different coding methods for Northeast dialect

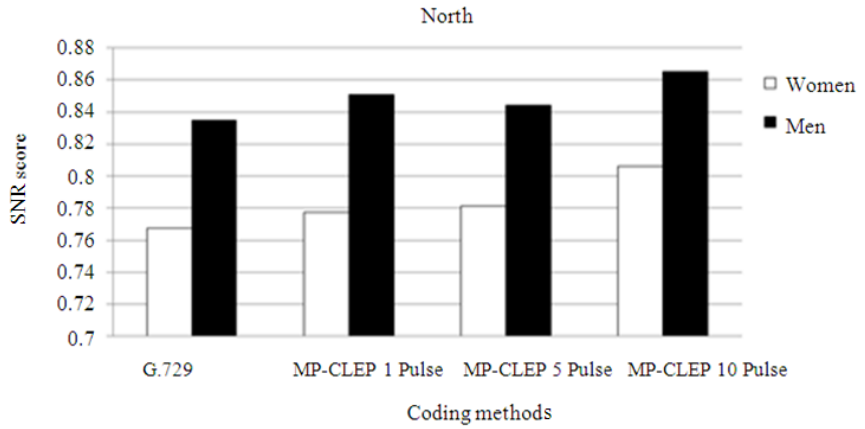

Fig. 10: Averaged SNR scores of different coding methods for north dialect

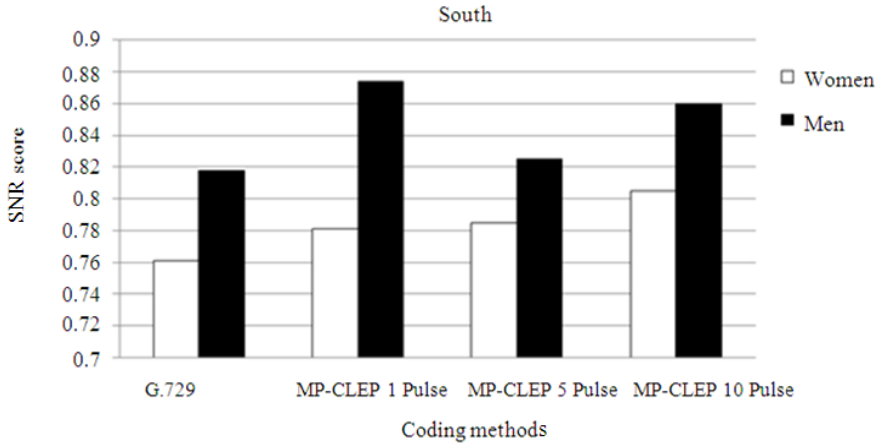

Fig. 11: Averaged SNR scores of different coding methods for south dialect

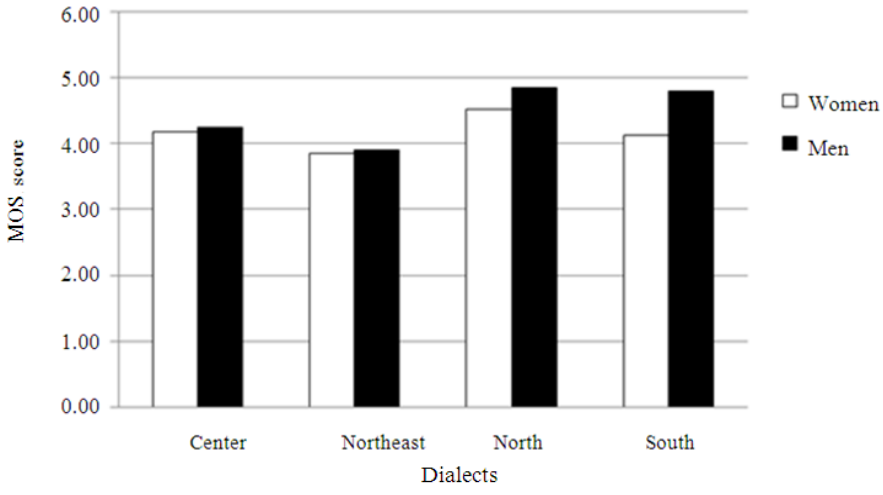

Fig. 12: Averaged MOS scores of all four dialects

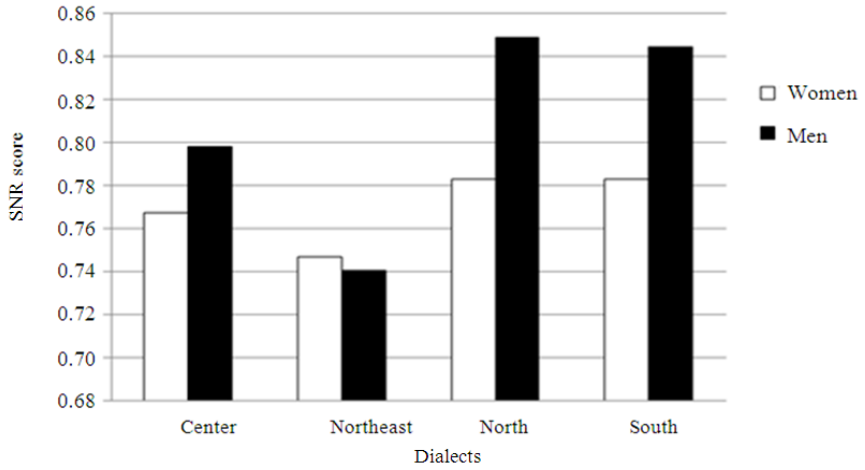

Fig. 13: Averaged SNR scores of all four dialects

DISCUSSION

From the subjective test results, considering the MOS scores in Fig. 12, the coding quality of North dialect is the highest level. The coding quality of South dialect is the second highest level, meanwhile the coding quality of Northeast dialect is the lowest level. Moreover, the coding quality of male speech is mostly higher than that of female speech. From the objective test results, considering the SNR scores in Fig. 13, the coding quality of all dialects corresponds to that of MOS scores in Fig. 12.

Furthermore, comparison of the coding quality among different coding methods has been conducted. From Fig. 4-7, the MP-CELP coder with 10 pulses in fixed codebook gives the best coding quality, while the CS-ACELP (G.729) coder gives the worst coding quality. From Fig. 8-11, the coding quality of all dialects corresponds to that of MOS scores in Fig. 4-7.

CONCLUSION

In this study, a comparative study of the coding quality of four main Thai dialects by using different low-bit-rate speech coders of the CS-ACELP coder and the MP-CELP coder has been conducted. The objective and subjective tests are used to evaluate the coding quality of four main dialects. Both tests show that the coding quality of North dialect is highest, meanwhile the coding quality of Northeast dialect is lowest. Moreover, the coding quality of male speech is mostly higher than that of female speech. From the study, it can be seen that the coding quality of all Thai dialects are different.

ACKNOWLEDGEMENT

The researcher is grateful to Kasetsart University for the research scholarship through the Center for Advanced Studies in Industrial Technology.

REFERENCES

Chomphan, 2010a. Performance evaluation of multi-pulse based code excited linear predictive speech coder with bitrates scalable tool over additive white Gaussian noise and Rayleigh fading channels. J. Comput. Sci., 6: 1433-1437. DOI: 10.3844/jcssp.2010.1438.1442

Chomphan, S., 2010b. Multi-Pulse based code excited linear predictive speech coder with fine granularity scalability for tonal language. J. Comput. Sci., 6: 1267-1271. DOI: 10.3844/jcssp.2010.1288.1292

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, May 10-12, 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 IEEE International Conference on Acoustics, Speech and Signal Processing, Apr. 14-17, IEEE Xplore Press, Toronto, Ont., 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, May 12-15, IEEE Xplore Press, Seattle, WA., 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, WA., pp: 153-156. DOI: 10.1109/ICASSP.1998.674390

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

Schroder, G. and M.H. Sherif, 1997. The road to G.729: ITU 8-kb/s speech coding algorithm with wireline quality. IEEE Commun. Mag., 35: 48-54. DOI: 10.1109/35.620525

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