

## Thai Speech Coding Based On Conjugate-Structure Algebraic Code Excited Linear Prediction Algorithm

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 mobile communication, speech coding aims at compressing the speech with lowest bitrate and highest quality for standard languages such as English, German and French. As for other languages with different uttering styles, the encoded speech quality is not guaranteed at the same bitrate. The appropriate evaluation should be performed to develop the speech quality by applying some suitable techniques. **Approach:** This study presents the comparison results of speech quality that is encoded and decoded by CS-ACELP coder according to ITU-G.729 standard. The purpose is to test the performance of CS-ACELP coder between Thai speech and English speech. **Results:** The study used 2 coding methods; (1) CS-ACELP coder without Voice Activity Detection and (2) CS-ACELP coder with Voice Activity Detection. The objective test was used to measure the speech quality for each case. The results show that both methods give Thai speech quality mostly below than English speech quality, as for methods comparison; both Thai and English, method (2) gives speech quality better than method (1). Eventually, we modified the coder by increasing the order of LP analysis to improve the Thai speech quality. **Conclusion:** From the finding, by no other modification, the quality of Thai coding is not equivalent to the English Language. After modifying the LP analysis by increasing the LP order from 10-12 or 14, the quality of Thai speech coding are truly improved. But the coding rate also increased for allocating the higher order information.

**Key words:** Thai speech, speech coding, Code-Excited Linear Predictive (CELP), International Telecommunications Union-Telecommunications Sector (ITU-T), Linear Predictive Coding (LPC), Voice Activity Detection (VAD), Line Spectrum Pairs (LSP)

---

### INTRODUCTION

Nowadays, the digital communications are widely developed. The information; audio, images, video or data can be transmitted through wire or wireless network channels (Jabrane *et al.*, 2007). Simultaneously, the number of users to access these networks increases rapidly. Consequently, channel capacity has to be increased, signal compression aims to perform this. As for speech, speech coding was created almost 60 years ago and improved from then until now (Cox, 1997; Hergli *et al.*, 2005).

The International Telecommunications Union-Telecommunications Sector (ITU-T) has already standardized 64 kb sec<sup>-1</sup>  $\mu$ A-law PCM, 32 kb sec<sup>-1</sup> ADPCM and 16 kb sec<sup>-1</sup> Low-Delay Code-Excited Linear Prediction (LD-CELP). The next step in the progression is an 8 kb sec<sup>-1</sup> speech coding algorithm. Since the three existing standards all provide high-quality and short-delay coding, the main requirement for the 8 kb sec<sup>-1</sup> algorithm is also initially high-quality

and short-delay coding (less than 5 ms) (Juan *et al.*, 1998; Chitra and Ravichandran, 2007).

To achieve high quality and short-delay coding at 8 kb sec<sup>-1</sup>, the backward adaptation technique was performed. Many coders used the Linear Predictive Coding (LPC) predictor in a backward-adaptive manner by performing LPC analysis on previously quantized speech. Since the reconstructed signal is available in both the encoder and decoder, this approach does not require that the LPC coefficients be sent to the decoder. However, although this technique is useful for 16 kb sec<sup>-1</sup> LD-CELP, an 8 kb sec<sup>-1</sup> coder was not found to give high quality when using only backward PLC analysis without pitch prediction. In 1995, CS-ACELP coder was developed and standardized as 8 kb sec<sup>-1</sup> G.729 (Chompun *et al.*, 2000; 2001a; 2001b; Chompun, 2004; Aqel, 2006; Moussaoui *et al.*, 2006).

To verify this circumstance, this study will study and compare the performance of CS-ACELP coder according to ITU G.729 between Thai and English speeches. The study will use 2 coding methods: the first is CS-ACELP coder without Voice Activity Detection

(VAD) and the second is CS-ACELP coder with Voice Activity Detection. The objective test as segmental signal to noise ratio (segSNR) will be selected to measure the speech quality for each case.

However, since Thai language is a tonal language. So the use of CS-ACELP coder following to ITU G.729 to compress Thai speech will be not guarantee the same quality as English language.

Later, we modified some parameters in CS-ACELP coder in order to improve its efficient for Thai speech compression. In the 10th linear prediction analysis and filtering, the order is increased to 12 and 14, respectively (Nomura *et al.*, 1998; Ozawa and Serizawa, 1998; Ozawa *et al.*, 1996).

### 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 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.

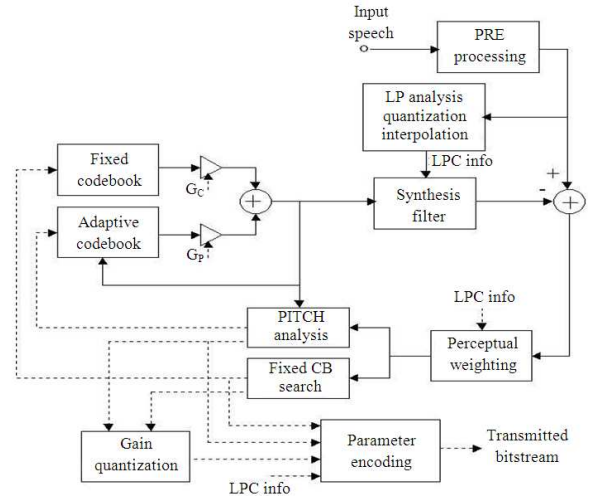


Fig. 1: Block diagram of G.729 encoder

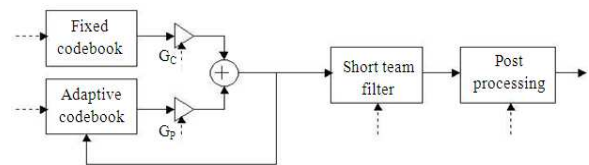


Fig. 2: Block diagram of G.729 decoder

The decoder principle is shown in Fig. 2. 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.

Table 1: First set of tested sentences (Chompun *et al.*, 2000; 2001a; Chompun, 2004; Sharmeela *et al.*, 2006)

Tested sentences (Thai)	Tested sentences (English)
ระบบป้องกันไฟต้องดี	This is a top secret.
อะแยงเข้าศึกษอยูบนเกาะเย	My name is John K. Smith.
เนื้อหมูเฝ้าพะและมึหนอน	Richard concluded the lecture.
ประโยคทดสอบภาษาไทย	Have a nice day for this trip.
กรุงเทพเป็นมหานคร	Leave me alone in the dark.
ดำเนินการได้ตามสะดวก	Wireless communications.

Table 2: Second set of tested sentences (Ounnpirak and Jitapunkul 1995; Ounnpirak, 1995).

Tested sentences (Thai)	Tested sentences (English)
อาจารย์เรียนเชิญผู้ปกครอง	Give me a zebra flag please.
ฉันฝากนกฮูกให้กับเธอ	Father played the xylophone.
ระซังถูกซ่อมแซมได้โกฏี	Bob quickly wore his slacks.
จพามีภกลงโทษหนัก	Jam did his paper last night.
เผ่าเก็บมณฑานฐานพระ	Excuse me, may I come in.
เมอเมเป็นยาร์กษาไรค	Good bye, see you tomorrow.

**Methods of coding:** In the experiment, 2 coding methods were used: One was CS-ACELP coder without VAD, another was CS-ACELP coder with VAD (Chomphan, 2010a; Chomphan, 2010b). Two sets of sentences were used by those coding methods. Then these two sets of sentences were compared. Each set contained 6 Thai and 6 English sentences. The first set of sentences was in Table 1, the second was in Table 2. In both sets, speeches of 3 men and 3 women were recorded.

The sentences chosen in these sets covered all of the characters in each language. The second set was performed as same as the first set to compare the results from each one.

## RESULTS

The quality of speech was evaluated in both 2 sets by using the values of segmental signal to noise ratio defined in Eq. 1 (Chompun *et al.*, 2000; 2001b; Chompun, 2004; Sharmeela *et al.*, 2006). Figure 3 and 4 show the original signal, the reconstructed signals of both methods of the first sentence in the first set for Thai and English respectively, while Table 3 shows segSNR for both methods:

$$\text{SegSNR} = \frac{10}{L} \sum_{i=0}^{L-1} \log_{10} \left\{ \frac{\sum_{n=0}^{N-1} s^2(iN+n)}{\sum_{n=0}^{N-1} (s(iN+n) - \hat{s}(iN+n))^2} \right\} \quad (1)$$

Where:

- N = Length of segment in bit
- L = Number of segments
- s(n) = The original signal
- $\hat{s}(n)$  = The reconstructed signal

Table 3: segSNR of Thai and English speech compression at LP order 10

		SegSNR(dB)			
		Thai		English	
Method		Male	Female	Male	Female
1st set	CS-ACELP	9.45	9.38	9.72	9.72
	CS-ACELP with VAD	9.50	9.42	9.75	9.73
2nd set	CS-ACELP	9.47	9.40	9.75	9.71
	CS-ACELP with VAD	9.49	9.44	9.76	9.72

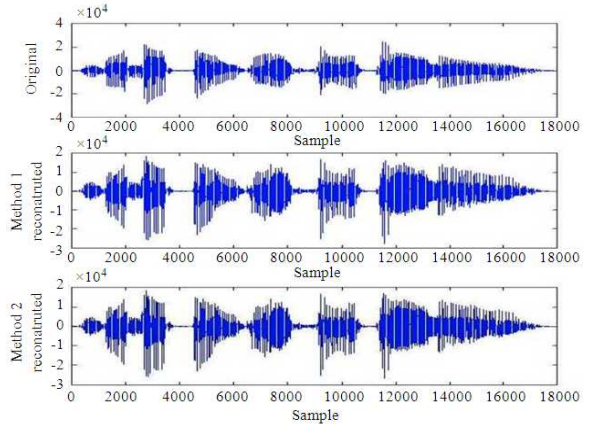


Fig. 3: Original signal, reconstructed signals of both methods. (Thai in the first set)

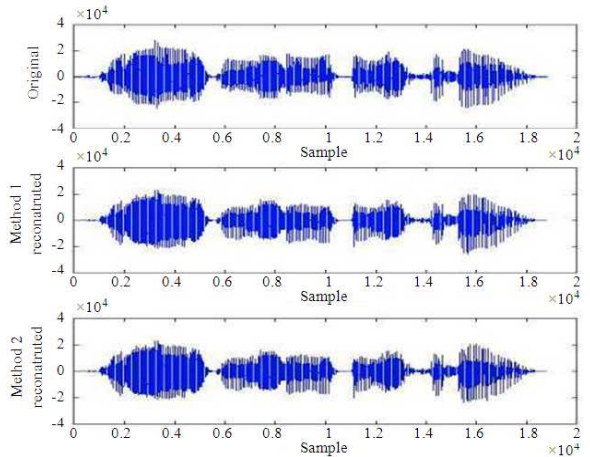


Fig. 4: Original signal, reconstructed signals of both methods. (English in the first set)

## DISCUSSION

The results in Table 3 show that both methods give Thai speech quality mostly below than English speech quality about 0.25-0.34 dB for the first set and about 0.27-0.31 dB for the second set. As for methods comparison; both Thai and English, method 2 gives speech quality better than method 1 about 0.01-0.05 dB for the first set and about 0.01-0.04 dB for the second set. Comparing 2 sets, the results of them were corresponding.

Table 4: segSNR of Thai speech compression at LP

Method		SegSNR(dB) - Thai					
		Order 10		Order 12		Order 14	
		Male	Female	Male	Female	Male	Female
1st set	CS-ACELP	9.45	9.38	9.53	9.45	9.54	9.48
	CS-ACELP with VAD	9.50	9.42	9.62	9.50	9.64	9.52
2nd set	CS-ACELP	9.47	9.40	9.54	9.45	9.53	9.48
	CS-ACELP with VAD	9.49	9.44	9.60	9.49	9.62	9.51

Finally, the order of LP analysis was increased to 12 and 14 to improve the quality of reconstructed speech in case of Thai speech. Table 4 shows the segSNR of both methods. For order 12, it shows the improvement of speech quality about 0.07-0.12 dB for the first set and about 0.05-0.11 dB for the second set, in comparison to LP order 10. For order 14, it shows the improvement of speech quality about 0.09-0.14 dB for the first set and about 0.06-0.13 dB for the second set, in comparison to LP order 10. Comparing 2 sets, the results of them were corresponding.

The results were shown that the quality of coding was improved. But the coding rate also increased for allocating the higher order information.

**CONCLUSION**

The ITU G.729 speech coder was applied to Thai Language and the performance evaluation was conducted. Without modifications, the quality of Thai coding is not equivalent to that of the English speech. After modifying the LP analysis by increasing the LP order from 10-12 or 14, the quality of Thai speech coding are explicitly improved. But the coding rate also increased for allocating the higher order information.

**ACKNOWLEDGEMENT**

The researcher wishes to thank Digital Signal Processing Research Laboratory, Faculty of Engineering, Chulalongkorn University, for the facility and technical support for this study and Kasetsart University at Si Racha campus for the research scholarship through the board of research.

**REFERENCES**

Aqel, M.J., 2006. An Approach for Developing Uncertain Multi-perspective Expert System Using the Algebraic Sum Method. *Am. J. Applied Sci.*, 3: 1719-1721. DOI: 10.3844/2006.1719.1721

Chitra, K. and V.C. Ravichandran, 2007. Turbo-coding of Coherence Multiplexed Optical PPM CDMA System with Balanced Detection. *Am. J. Applied Sci.*, 4: 264-268. DOI: 10.3844/2007.264.268

Chomphan, 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, 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: 1433-1437. DOI: 10.3844/jcssp.2010.1433.1437

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, Tainan, Taiwan*, pp: 197-200. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=1412726](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1412726)

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, IEEE, Mar. 27-29, Las Vegas, USA.*, pp: 461-464. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=918839](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=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 Systemic, Cybernetics and Informatics, July 22-25, IIC, Orlando, Florida, USA.*, pp: 519-522. [http://daisy.ee.eng.chula.ac.th/~d1oatty/oat\\_files/group\\_files/study/SCI2001\\_S\\_Chompun.pdf](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>
- Cox, R.V., 1997. Three new speech coders from the ITU cover a range of applications. *IEEE Commun. Mag.*, 35: 40-47. DOI: 10.1109/35.620524
- Hergli, M., J. Baili, F. Bousslama and K. Besbes, 2005. A New Compressing Ultrasonic Data Algorithm Based on Wavelets. *Am. J. Applied Sci.*, 2: 1615-1618. DOI: 10.3844/2005.1615.1618
- 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
- Juan, L., L. Biqin and F. Qiuliang, 1998. An 8 kb sec<sup>-1</sup> conjugate-structure algebraic CELP (CS-ACELP) speech coding, Proceeding of the IEEE International Conference on Signal Processing, IEEE, Oct. 12-16, Beijing, China, pp: 1729-1732. DOI: 10.1109/ICOSP.1998.7709
- Moussaoui, A., Y. Selaimia and H.A. Abbassi, 2006. Hybrid Hot Strip Rolling Force Prediction using a Bayesian Trained Artificial Neural Network and Analytical Models. *Am. J. Applied Sci.*, 3: 1885-1889. DOI: 10.3844/2006.1885.1889
- 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, IEEE, May 12-15, Seattle, USA., pp: 341-344. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=674437](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=674437)
- Ozawa, K. and M. Serizawa, 1998. High quality multi-pulse based CELP speech coding at 6.4 kb sec<sup>-1</sup> and its subjective evaluation. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, May 12-15, Seattle, USA., pp: 529-532. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=674390](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=674390)
- Ozawa, K., T. Nomura, M. Serizawa, 1996. MP-CELP speech coding based on multi-pulse vector quantization and fast search. *IEICE Trans.*, 79: 1655-1663. 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 sec<sup>-1</sup> speech coding algorithm with wireline quality. *IEEE Commun. Mag.*, 35: 48-54. DOI: 10.1109/35.620525
- Sharmeela, C., M.R. Mohan, G. Uma and J. Baskaran, 2006. A Novel Detection and Classification Algorithm for Power Quality Disturbances using Wavelets. *Am. J. Applied Sci.*, 3: 2049-2053. DOI: 10.3844/2006.2049.2053