Journal of Computer Science 7 (6): 865-870, 2011
ISSN 1549-3636
© 2011 Science Publications

Analysis of Fundamental Frequency Contour of Coded Speech Based on Multi-Pulse Based 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 low-bit-rate speech communication, speech coding deteriorates the characteristics of the coded speech significantly. An important feature of the speech is the fundamental frequency contour which determines the pitch information of the speech. It has been known that pitch information is one of the core parameter of the multi-pulse based code excited linear prediction (MP-CELP) speech coder. Therefore the study of the deteriorated fundamental frequency contour should be conducted properly. **Approach:** This study proposes an analysis of the fundamental frequency contour of the coded speech based on MP-CELP speech coder. The comparison of the fundamental frequency contour of the natural speech and that of the coded speech has been performed. The MP-CELP with three levels of bitrate scalability is selected as the core speech coder. The speech material includes a hundred of male speech utterances and a hundred of female speech utterances. **Results:** The experimental results show that the speech coder causes the deterioration of the fundamental frequency contour empirically. The Root Mean Square Error (RMSE) between the fundamental frequency contour of the natural speech and that of the coded speech for three different bitrates has been conducted. The lower bitrate causes the higher value of RMSE. **Conclusion:** From the study, it is a proved that the MP-CELP speech coder deteriorates the fundamental frequency contour of the transmitted speech.

Key words: Multi-Pulse based Code Excited Linear Predictive (MP-CELP), speech coding, bitrate scalability, Linear Prediction (LP), fundamental frequency contour, line spectrum pairs

INTRODUCTION

In speech communication network with an increasing number of users, channel capacity is needed to be increased, 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. These kinds of applications require special considerations for packet loss. To treat this problem, a bitrate-scalable coder is developed 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 has been conducted at the MPEG-4 (Nomura *et al*., 1998; Sen, 2005; Sen *et al*., 2005; Chomphan, 2010a; 2010b). 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. Subsequently, MP-CELP coder has been proposed to be a scalable coder around this bitrate. With the flexibility attribute, 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 (Nomura *et al*., 1998; Melesse and Hanley, 2005; Chomphan, 2010b; Sathyabalan *et al*., 2009).

In low-bit-rate speech compression in the communication system, the coding quality should be concerned in various aspects including bitrate, coded speech quality, coding complexity and coding delay. The coding speech quality is one of the most important issues that should be considered. The naturalness and the intelligibility of the coded speech are two main points for developing the speech coder. We, consequently, take into account of the fundamental frequency contour which is implicitly related to the pitch of the speech; the main parameter of the generation of the excitation signal of the MP-CELP speech coder (Chomphan, 2010a; 2010b).

$$\overset{T}{C_i(C_i)\ V(V)C_f}$$

Fig. 1: Thai syllable structure

The fundamental frequency contour plays a vital role on the naturalness and the intelligibility of the coded speech. High accuracy of the fundamental frequency brings about the naturalness of the speech and also the correctness of the tonal feature which is implicitly related to the intelligibility of the speech. Therefore, it is needed to explore the fundamental frequency in the study of the speech coder with low bitrate.

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; Ghaderi *et al*., 2005; Rahman *et al*., 2009). For tonal language, such as Thai, a syllable is composed of consonants, vowels and tone (Wutiwiwatchai and Furui, 2007). 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, e.g., (Chompun *et al*., 2000; Wutiwiwatchai and Furui, 2007). Therefore, the bitrate scalable tonal language speech coder based on a multi-pulse based code excited linear predictive coding (Taumi *et al*., 1996; Ozawa *et al*., 1996) has been proposed. Moreover, this coder is improved for the tonal language speech by applying the high pitch delay resolutions to retain the tone information precision.

This study proposes an analysis of the fundamental frequency contour of the coded speech based on MP-CELP speech coder with the high pitch delay resolutions for tonal speech. The comparison of the fundamental frequency contour of the natural speech and that of the coded speech has been performed. The MP-CELP with three levels of bitrate scalability is selected as the core speech coder. The speech material includes a hundred of male speech utterances and a hundred of female speech utterances.

MATERIALS AND METHODS

Bitrate scalable MP-CELP coder: 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. 2 (Taumi *et al*., 1996; 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. 3. 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.

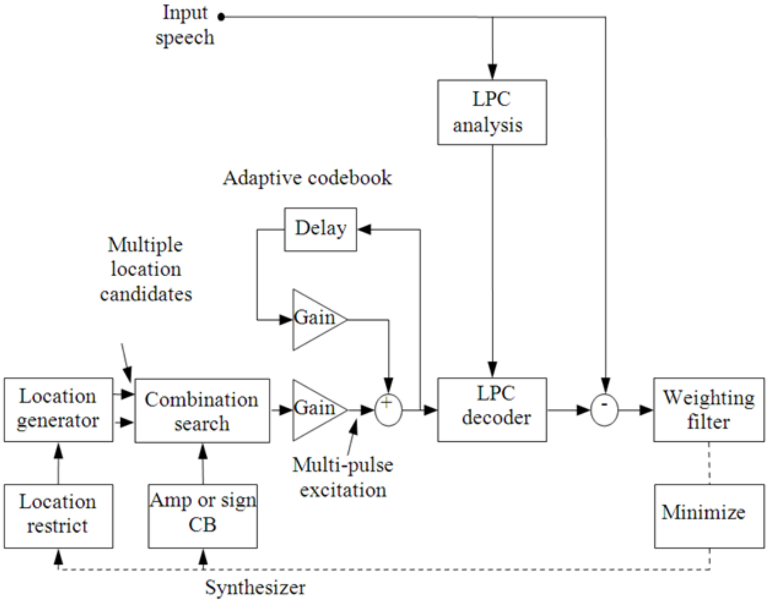


Fig. 2: MP-CELP core coder

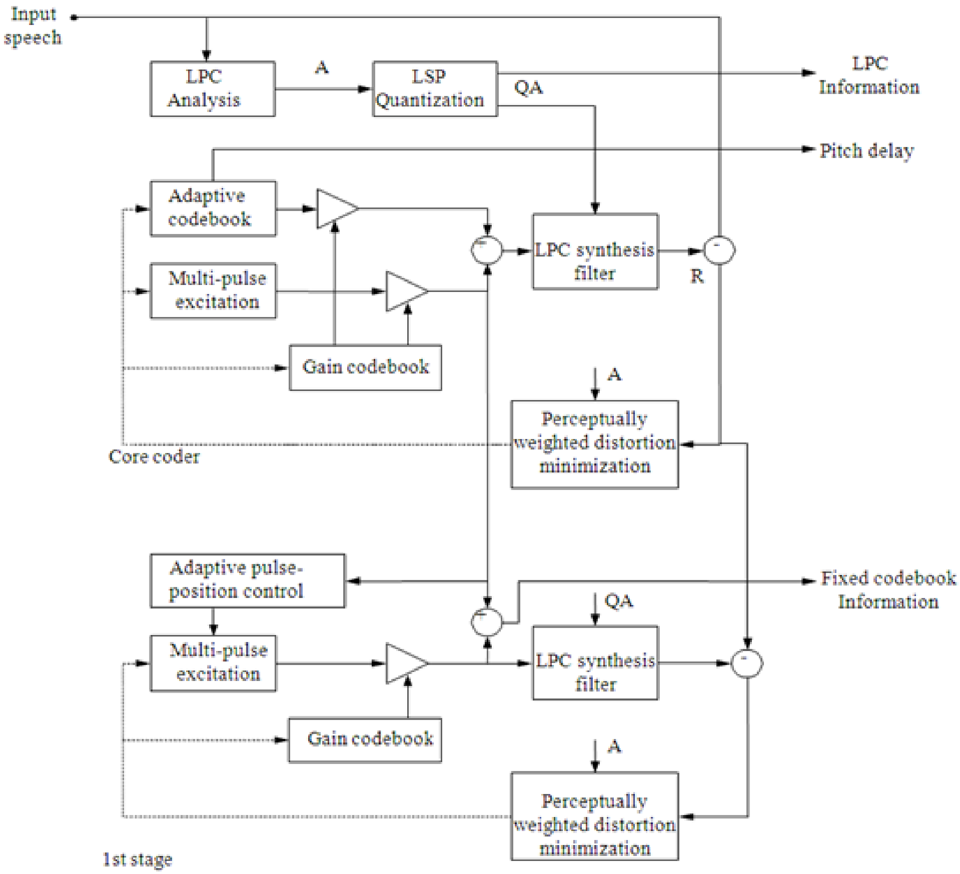


Fig. 3: One-stage bitrate scalable MP-CELP coder

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

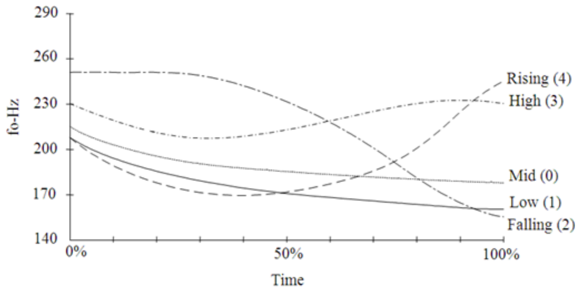


Fig. 4: f0 characteristic of 5 tones in Thai

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.

Tonal language speech coder : 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*., 2000; Wutiwiwatchai and Furui, 2007). 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 (or f0) significantly involves in tone of tonal language, this study proposes an improvement of the bitrate scalable MP-CELP coder by applying the High Pitch Delay Resolutions (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.

Fundamental frequency contour (F0 contour): There is a substantial amount of data on the frequency of the voice fundamental or fundamental frequency (F0) in the speech of speakers who differ in age and sex. The data have been published for several languages and for various types of discourse. The data always include an average measure of F0, usually expressed in Hz, but in some cases the average duration of a period has been reported instead. Typical values obtained for F0 are 120 Hz for male speech and 210 Hz for female speech (Waldstein and Boothroyd, 1994; Khor *et al*., 2009). An example of F0 contour of the natural speech is depicted in Fig. 1.

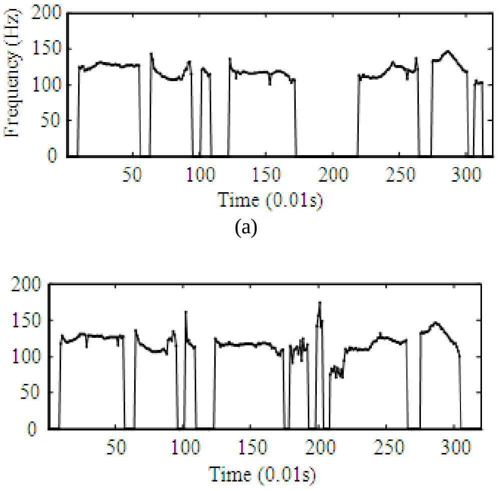
Typically, the mean values of F0 change slightly with age. For female speech, F0 is quite stationary up to the period of menopause, when it decreases to reach the minimum which is about 15 Hz lower around 70 years of age (Pegoraro-Krook, 1988). The physiological changes is an effect of the increased testosterone-oestrogen ratio at that period. A similar decreasing of F0 can be caused by the habit of smoking (Gilbert and Weismer, 1974; Rushaidin *et al*., 2009). For male speech, the dramatic decrease in F0 during puberty duration has been observed to continue with subsequent deceleration until about 35 years of age. Thereafter, at about 55 years of age, F0 begins to rise again (Pegoraro-Krook, 1988).

RESULTS

To analyze of the fundamental frequency contour of the coded speech based on MP-CELP speech coder with the high pitch delay resolutions for tonal speech. The comparison of the fundamental frequency contour of the natural speech and that of the coded speech has been performed. An example of the comparison pair of the fundamental frequency contours from them is illustrated in Fig. 5. The MP-CELP with three levels of bitrate scalability is selected as the core speech coder. The selected bitrates are 5600, 8200 and 12200 bps.

Table 2: Examples of Thai tested utterances 0, 1, 2, 3, 4 at each syllable represents tone number

Order	Tested utterances
1	การประชุมทางวิชาการ งที่ครั้ 1 k-aa-n^-0|pr-a-z^-1|ch-u-m^-0|th-aa-ng^-0|w-i-z^-3|ch-aa-z^-0|k-aa-n^-0|khr-a-ng^-3|th-ii-z^-2|n-v-ng^-1|
2	โครงการวิจัยและพัฒนาอิเล็กทรอนิกส์และคอมพิวเตอร์ khr-oo-ng^-0|k-aa-n^-0|w-i-z^-3|c-a-j^-0|l-x-z^-3|ph-a-t^-3|th-a-z^-3|n-aa-z^-0|z-i-z^-1|l-e-k^-3|thr-@@-z^-0|n-i-k^-1|l-x-z^-3|kh-@-m^-0|ph-i-w^-3|t-qq-z^-2|
3	ปีงบประมาณ 2531 p-ii-z^-0|ng-o-p^-3|pr-a-z^-1|m-aa-n^-0|$-$-$-$|s-@@-ng^-4|ph-a-n^-0|h-aa-z^-2|r-@@-j^-3|s-aa-m^-4|s-i-p^-1|z-e-t^-1|
4	กระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน k-a-z^-1|r-a-z^-1|th-a-z^-3|r-uua-ng^-0|w-i-t^-3|th-a-z^-3|j-aa-z^-0|s-aa-z^-4|s-o-t^-1|th-e-k^-3|n-oo-z^-0|l-oo-z^-0|j-ii-z^-0|$-$-$-$|l-x-z^-3|k-aa-n^-0|ph-a-z^-3|l-a-ng^-0|ng-aa-n^-0|



(a)

(b)

Fig. 5: An example of the comparison pair of the fundamental frequency contours from the natural speech and that of the coded speech

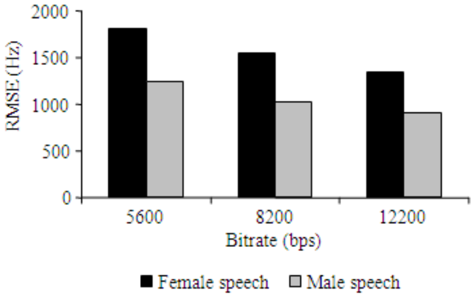


Fig. 6: Root mean square errors between fundamental frequency contours of the natural speech and that of the coded female and male speech at different bitrates

The speech material includes a hundred of male speech utterances and a hundred of female speech utterances. Four examples of Thai tested utterances are represented in Table 2. The root mean square errors between fundamental frequency contours of the natural speech and that of the coded speech are calculated as shown in Fig. 6.

DISCUSSION

From Fig. 5, the comparison of the fundamental frequency contours from the natural speech and that of the coded speech shows that both there are some empirical differences which cause the deterioration in naturalness and intelligibilities of the coded speech. In other words, the characteristics of the natural speech are changed by the coding algorithm. From Fig. 6, it has been observed that most of the RMSE of female speech are higher than that of male speech. Moreover, it has been seen that the value of RMSE depends on the coding rates. The coding rate of 5600 bps gives the highest values of RMSEs for both male and female speech, while the coding rate of 12200 bps gives the lowest values of RMSEs for both male and female speech.

CONCLUSION

This study proposes an analysis of the fundamental frequency contour of the coded speech based on MP-CELP speech coder at three core coding rates of 5600, 8200 and 12200 bps. The speech material covers a hundred of male speech utterances and a hundred of female speech utterances. The results show that the speech coder causes the deterioration of the fundamental frequency contour empirically. Moreover, the root mean square error between the fundamental frequency contour of the natural speech and that of the coded speech for three different bitrates has been conducted. The lower bitrate causes the higher value of RMSE. From the study, it has been confirmed that the MP-CELP speech coder deteriorates the fundamental frequency contour of the transmitted speech.

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: 1433-1437. 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, May 10-12, NECTEC, Chiangmai, Thailand, pp: 1-5. http://daisy.ee.eng.chula.ac.th/~d1oatty/oat_files/group_files/study/SNLP2000_Supattarachai_final.pdf

Ghaderi, S.F., M.A. Azadeh and S. Bamdad, 2005. Analyzing the electricity consumption using experimental design technique. Am. J. Applied Sci., 2: 1464-1470. DOI: 10.3844/.2005.1464.1470

Gilbert, H.R. and G.G. Weismer, 1974. The effects of smoking on the speaking fundamental frequency of adult women. J. Psychol. Res., 3: 225-231. DOI: 10.1007/BF01069239

Khor, S.F., Z.A. Talib, H.A.A. Sidek, W.M. Daud and B. H. Ng, 2009. Effects of ZnO on dielectric properties and electrical conductivity of ternary zinc magnesium phosphate glasses. Am. J. Applied Sci., 6: 1010-1014. DOI: 10.3844/ajassp.2009.1010.1014

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

Melesse, A.M. and R.S. Hanley, 2005. Energy and carbon flux coupling: Multi-ecosystem comparisons using artificial neural network. Am. J. Applied Sci., 2: 491-495. DOI: 10.3844/.2005.491.495

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, Seattle, USA., pp: 341-344. 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/s and its subjective evaluation. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 12-15, IEEE, Seattle, USA., pp: 529-532. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=674390

Ozawa, K., T. Nomura and 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

Pegoraro-Krook, M. I., 1988. Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis. Folia Phoniatrica, 40: 82-90. DOI: 10.1159/000265888

Rahman, M.M., A.K. Ariffin, S. Abdullah, M.M. Noor and R.A. Bakar *et al*., 2009. Assessment of surface treatment on fatigue life of cylinder block for linear engine using frequency response approach. Am. J. Applied Sci., 6: 715-725. DOI: 10.3844/ajassp.2009.715.725

Rushaidin, M.M., S.H. Salleh, T.T. Swee, J.M. Najeb and A. Arooj, 2009. Wave V detection using instantaneous energy of auditory brainstem response signal. Am. J. Applied Sci., 6: 1669-1674. DOI: 10.3844/ajassp.2009.1669.1674

Sathyabalan, P., V. Selladurai and P. Sakthivel, 2009. ANN based prediction of effect of reinforcements on abrasive wear loss and hardness in a hybrid MMC. Am. J. Eng. Applied Sci., 2: 50-53. DOI: 10.3844/ajeassp.2009.50.53

Sen, M.D.L., 2005. Asymptotic hyperstability of dynamic systems with point delays. Am. J. Applied Sci., 2: 1279-1282. DOI: 10.3844/.2005.1279.1282

Sen, M.D.L., J.L. Malaina, A. Gallego and J.C. Soto, 2005. Stability of non-neutral and neutral dynamic switched systems subject to internal delays. Am. J. Applied Sci., 2: 1481-1490. DOI: 10.3844/.2005.1481.1490

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

Waldstein, R.S. and A. Boothroyd, 1994. Speech reading enhancement using a sinusoidal substitute for voice fundamental frequency. Speech Commun., 14: 303-312. DOI: 10.1016/0167-6393(94)90024-8

Wutiwiwatchai, C. and S. Furui, 2007. Thai speech processing technology: A review. Speech Commun., 49: 8-27. DOI: 10.1016/j.specom.2006.10.004