Journal of Computer Science 7 (3): 379-386, 2011
ISSN 1549-3636

Analytical Study on Fundamental Frequency Contours of Thai Tones using Tone-Geometrical Model

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 tonal language speech; such as Thai, Mandarin and Vietnam, tone is an important feature of a syllable that must be taken into consideration, since tone is a supra-segmental feature in the speech prosody. Modeling of tone with high accuracy could enhance the quality of synthesized speech in the speech synthesis system. This study focuses on a model analysis of fundamental frequency (F0) contours of Thai tones using tone-geometrical model. **Approach:** Tone-geometrical model applied is this study is a basic model which is expected to extract the dimensional features of a syllable-length portion of fundamental frequency contour. Seven selected parameters are extracted from a syllable-length portion of fundamental frequency contour and then are analyzed. **Results:** In the experiments, 2,500 speech utterances from TSynC-1 speech database were selected and used as speech materials. The distributions for all seven parameters are presented. The statistical figures of mean and standard deviation values from five tones are also calculated. The results show that most of the proposed parameters can distinguish five Thai tones explicitly. **Conclusion:** From the finding, the proposed parameters of tone-geometrical model could be further applied in the speech or other speech processing technologies.

Key words: Thai tones, tone-geometrical model, fundamental frequency contours, analysis of fundamental frequency, parameter analysis, dynamic range, contour_slope, synthesis system, syllable levels, analytical study

INTRODUCTION

Tone analysis issue has been conducted in a number of speech technology research fields for years in many tonal speech languages. In speech processing area; including speech recognition, speech synthesis, speech analysis and speech coding, an appropriate tone modeling of a portion of F0 contour contributes the effectiveness of the implemented speech processing systems. The former study on F0 modeling has been considerably conducted in various speech units and several techniques such as utterance level (Fujisaki and Ohno, 1998; Fujisaki *et al*., 1990; Tao *et al*., 2006; Saito and Sakamoto, 2002; Ni and Hirose, 2006; Li *et al*., 2004), word and syllable levels (Fujisaki *et al*., 1990; Fujisaki and Sudo, 1971). In Thai speech, Fujisaki's model has been successfully applied for modeling of utterances, tones and words (Hiroya and Sumio, 2002; Seresangtakul and Takara, 2002; 2003). In the Thai speech synthesis, Chomphan and Kobayashi implemented a speaker-dependent and speaker-independent systems in 2007-2009 (Chomphan and Kobayashi, 2007; 2008; 2009), in which the F0 contour was modeled using statistical approach. Moreover, the speaker-independent system also used the Fujisaki's model in the extended modules. In another approach, tone-geometrical model is a simple geometrical-structure syllable-level model applied in a speech synthesis system (Chomphan and Kobayashi, 2009). However, the it has not been performed thoroughly to model all five Thai tones. Therefore this study proposed an analysis of F0 modeling of Thai tones using the tone-geometrical model which is a preliminary study for the advanced research in speech synthesis and recognition.

MATERIALS AND METHODS

Tone-geometrical model: The F0 contour is treated as a concatenation of a number of the sequential portion of tone-geometrical models as depicted in Fig. 1.

Seven selected parameters are calculated based on the following criteria:

Parameter 1: dur = t_final
Parameter 2: F0_init
Parameter 3: delta_F0 = F0_final - F0_init
Parameter 4: F0_range = F0_max - F0_min

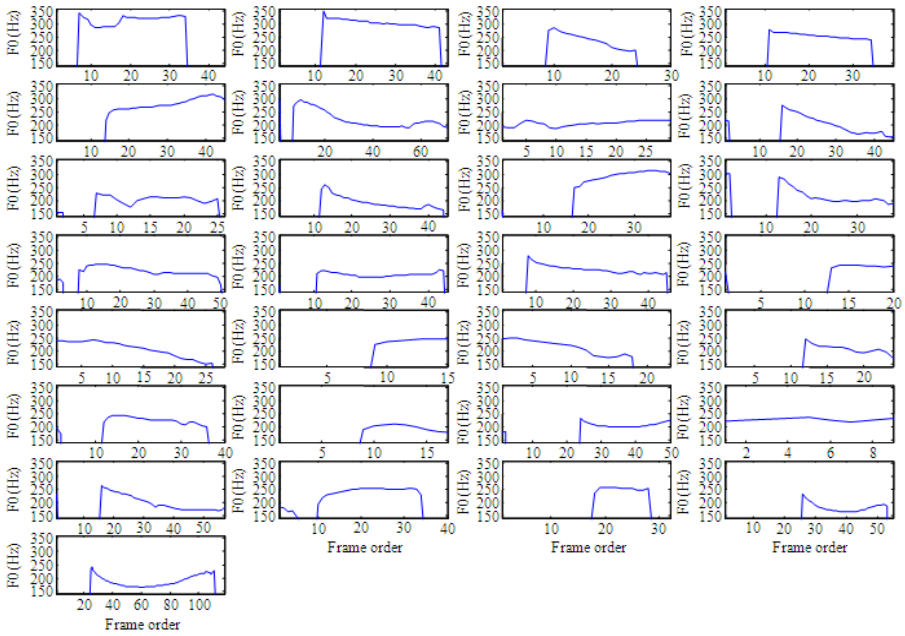

Fig. 1: The F0 potions of all syllables from an utterrance (1 frame = 5 ms)

Parameter 5: sign_F0_range = sign(delta_F0)*F0_range
Parameter 6: contour_slope = delta_F0 / dur
Parameter 7: sign_contour_slope = sign_F0_range/dur

Parameter 1 or dur denotes the syllable duration which can be obtained directly from t_final as shown in Fig. 2. Parameter 2 or F0_init is the initial value of a portion of an F0 contour as depicted in Fig. 2. Parameter 3 or delta_F0 is the difference between two frequencies of the final frequency and the initial frequency. Parameter 4 or F0_range represents the dynamic range of the portion in Hertz or the difference between two frequencies of the maximum frequency and the minimum frequency. Parameter 5 or sign_F0_range is the F0_range with a sign where the positive value represents the upward movement while the negative sign represents the downward movement. Parameter 6 or contour_slope is the ratio between delta_F0 and dur. This parameter reflects the gradient magnitude of the portion. Finally, parameter 7 or sign_contour_slope is the ratio between sign_F0_range and dur. This parameter reflects the gradient magnitude of the portion and also the direction movement. It has been noted that the sign_contour_slope applies the dynamic range of the portion meanwhile the contour_slope applies the delta_F0.

Procedures of parameter analysis: The following procedures of parameter analysis are implemented for an utterance from the speech data material:

- Extracting the F0 values from the speech raw file
- Extracting the beginning time and ending time of all syllables from the label file and converting them to the corresponding frame number
- Cutting the F0 intervals for all syllables from step 1 by using frame numbers in step 2
- Eliminating the interval with some non-sense and zero-value F0s from step 3
- Calculating the tone-geometrical model parameters of the F0 portion from step 4 by using the early definitions
- Plotting the distribution of parameters over its range
- Calculating the statistical values of the parameters from the distributions in step 6

It should be noted that the output of F0 potions from step 4 is little different from the output of F0 potions from step 3 as depicted in Fig. 3., since the interval of non-sense and zero-value F0s are absolutely eliminated. The term “non-sense F0” refers to the abnormal F0 value which is largely different from the neighboring F0s. In some cases, the “non-sense F0” means the F0s from the adjacent portion of F0 contour. These non-sense F0s can deteriorate the geometrical feature of the model, therefore they should be eliminated before calculating the model parameters. The term “zero-value F0” refers to the non-existing F0 which means that the F0 cannot be calculated. This non-existing F0 usually locates in the unvoiced or voiceless region of speech (Zainal *et al*., 2009).

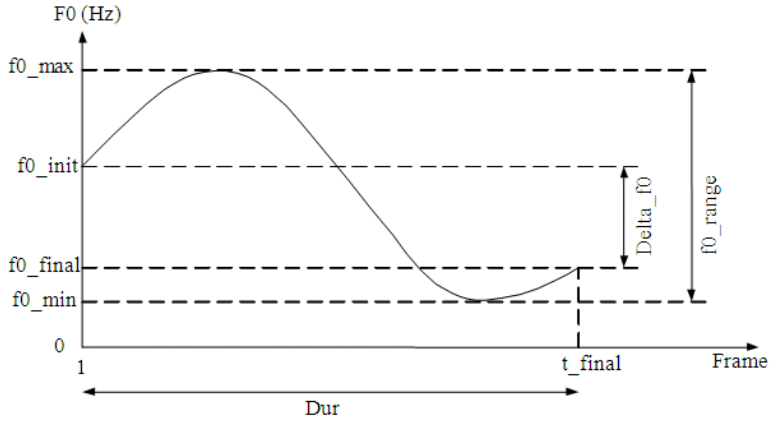

Fig. 2: Tone-geometrical model extracted from a portion of an F0 contour

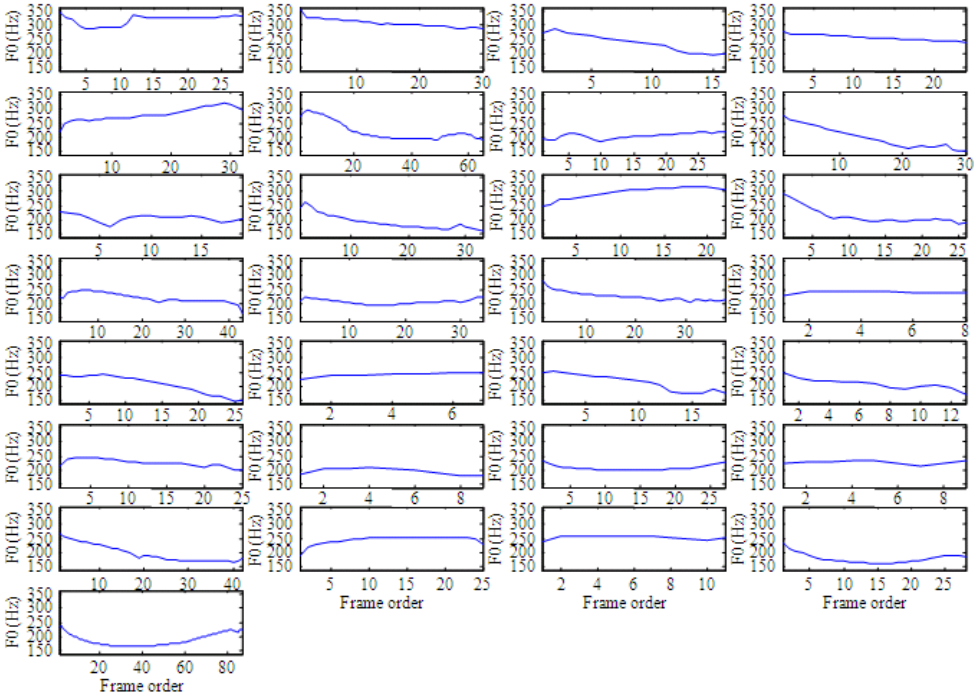

Fig. 3: The F0 potions of all syllables from an utterance with eliminating of nonsense, zero-value F0 values from step 4 of procedures of parameter analysis (1 frame = 5 ms)

RESULTS

The 2500 speech utterances from TSynC-1 speech corpus of NECTEC are used for model analyzing. (Chomphan, 2009; Alshamasin, 2009; Chomphan, 2010a; 2010b; 2010c; 2010d; 2010e).

In each of the selected parameter, we analyzed the frequency distribution over its range and then the distributions for all five tones in Thai are comparative shown in Fig. 4-10.

Figure 4-10 present the comparison among the distributions of the selected parameters including parameter 1-7, respectively for five tones in Thai and a combined tone ("alltone" notation) to show the differences and similarities among those tones. The "tone0", "tone1", "tone2", "tone3" and "tone4" denote middle tone, low tone, falling tone, high tone and rising tone, respectively.

From all of these frequency distribution graphs, the first and second statistical moments (mean and standard deviation values) were subsequently calculated and shown in Table 1.

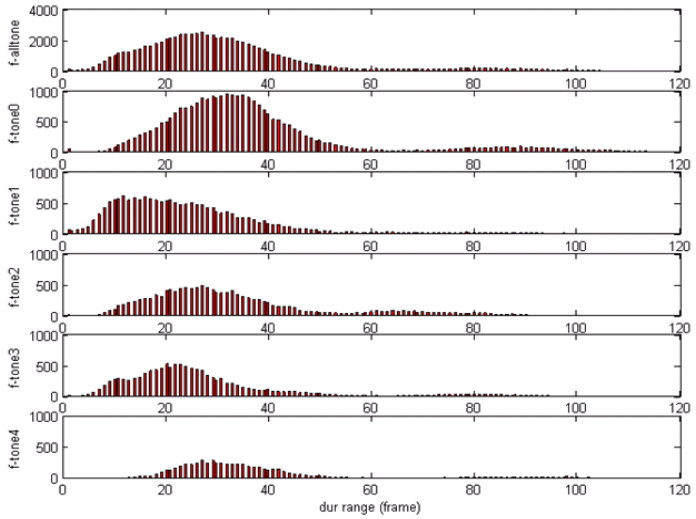

Fig. 4: Parameter 1: Distribution of dur parameter over its range (1frame = 5ms)

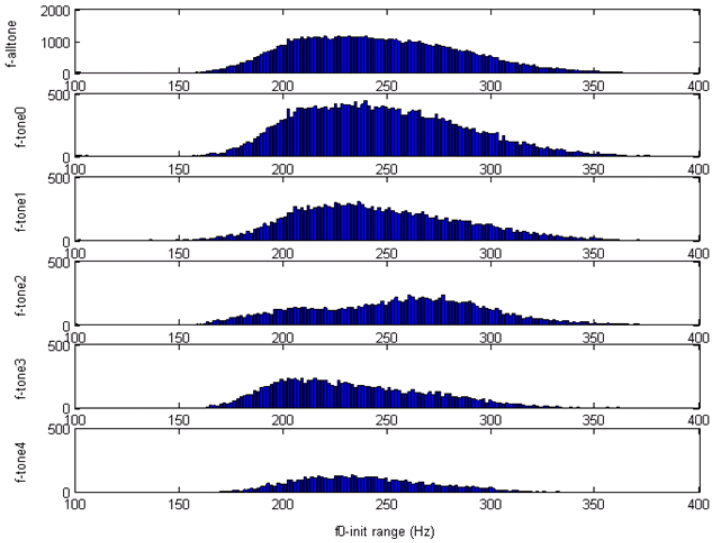

Fig. 5: Parameter 2: Distribution of F0_init parameter over its range

Table 1: Statistical figures of the seven selected sparameters

	Statistical figures	Alltone	Tone0	Tone1	Tone2	Tone3	Tone4
	Number of syllables	77413.00	26734.00	17650.00	13664.00	12656.00	6709.00
Mean	dur	32.99	37.58	25.81	33.96	29.04	39.37
	F0_init	243.30	245.01	243.52	251.99	232.60	238.37
	delta_F0	-22.24	-35.53	-48.77	13.53	10.21	-33.53
	F0_range	60.87	60.12	69.82	62.52	46.30	64.39
	sign_F0_range	-28.74	-46.01	-61.08	16.28	13.35	-45.97
	contour_slope	-0.79	-1.05	-2.25	0.70	0.35	-1.10
	sign_contour_slope	-0.99	-1.34	-2.78	0.90	0.47	-1.46
SD	dur	19.96	20.34	17.63	18.27	19.32	21.36
	F0_init	45.52	44.48	47.71	47.34	42.97	39.74
	delta_F0	48.70	38.60	42.05	51.40	36.84	39.55
	F0_range	32.30	30.07	32.47	34.60	18.96	24.76
	sign_F0_range	62.62	49.00	44.73	68.65	39.68	44.35
	contour_slope	2.05	0.60	2.38	1.16	1.11	0.43
	sign_contour_slope	2.52	0.73	2.71	1.62	1.44	0.57

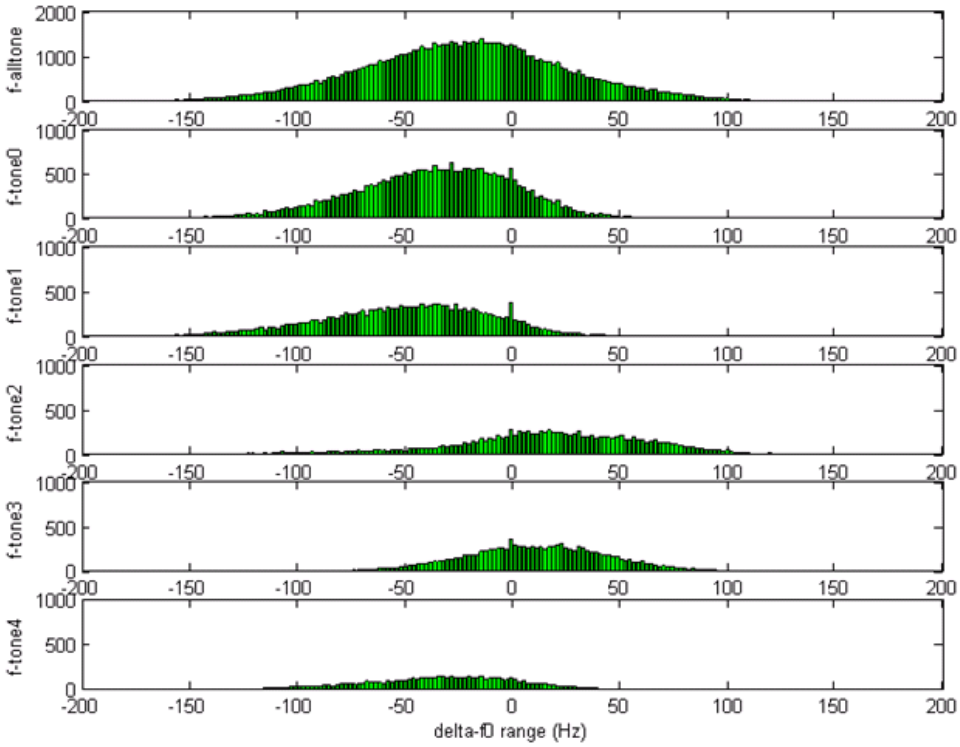

Fig. 6: Parameter 3: Distribution of delta_F0 parameter over its range

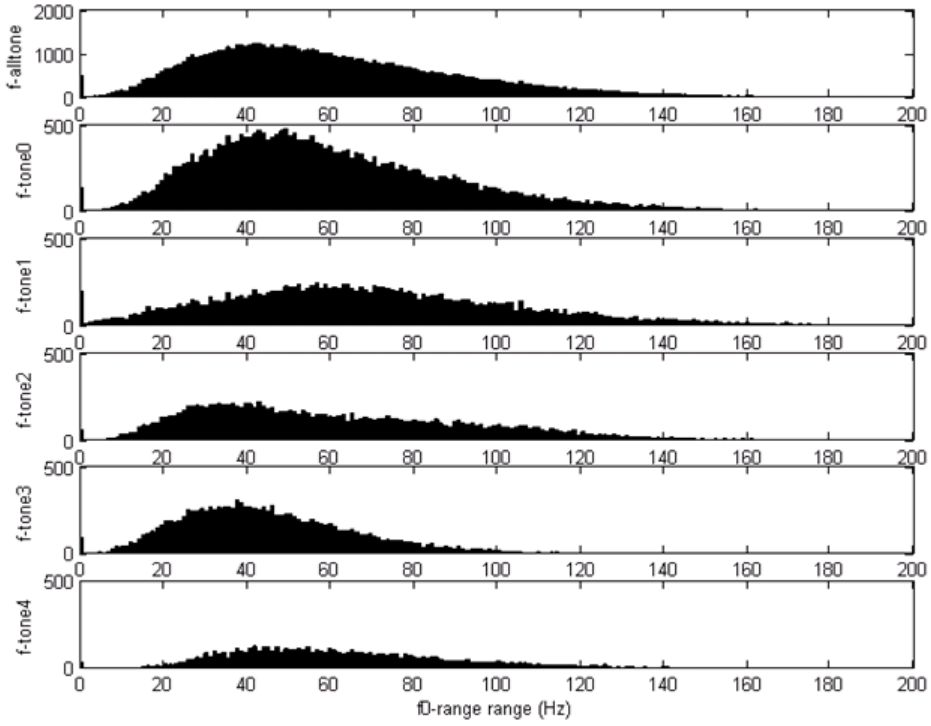

Fig. 7: Parameter 4: Distribution of F0_range parameter over its range

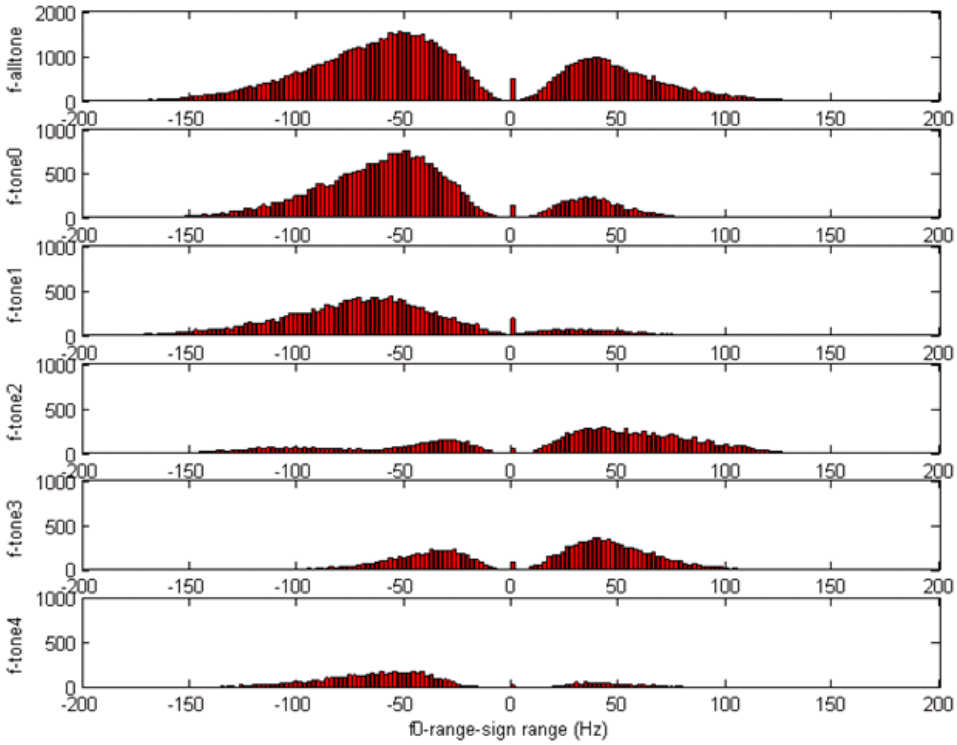

Fig. 8: Parameter 5: Distribution of sign_F0_range parameter over its range

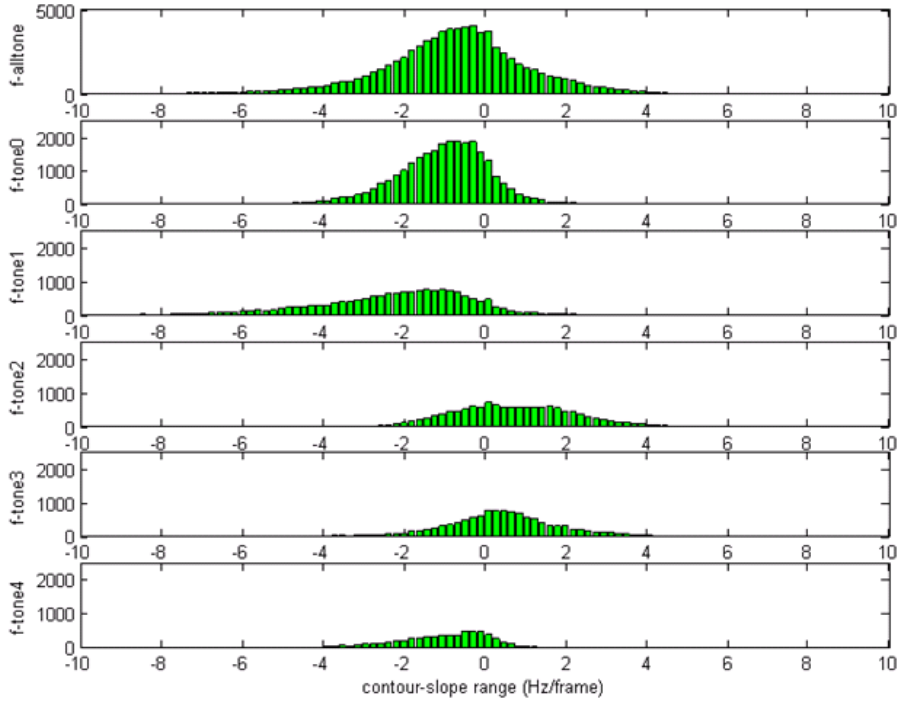

Fig. 9: Parameter 6: Distribution of contour_slope parameter over its range (1 Hertz/frame = 45°, 10 Hertz/frame = 84.3°)

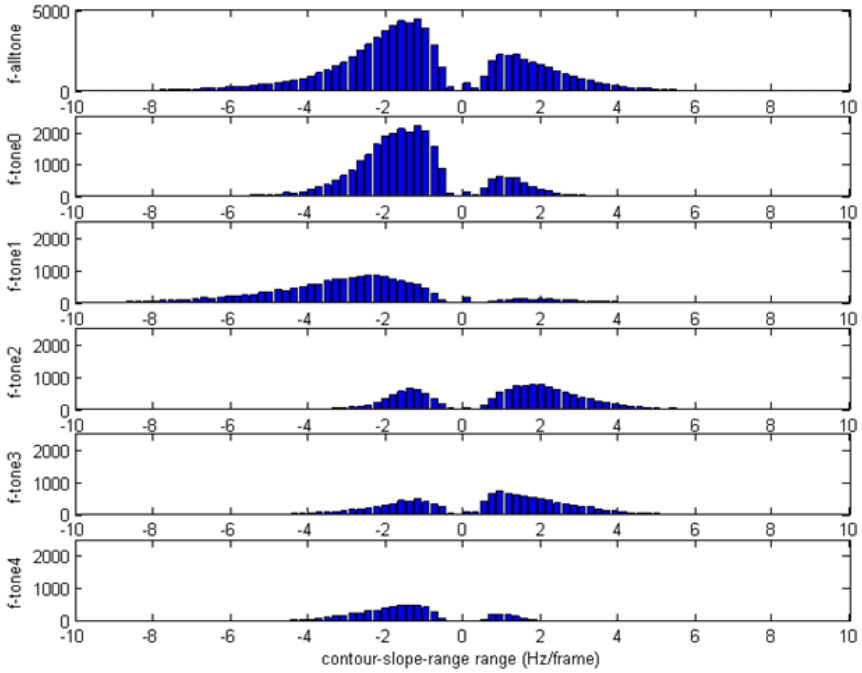

Fig. 10: Parameter 7: Distribution of sign_contour_slope parameter over its range (1 Hertz/frame = 45°, 10 Hertz/frame = 84.3°)

DISCUSSION

From the frequency distribution graphs in Fig. 4-10, most results show that the five distributions of each tones are significantly different. Except for only some cases, the distributions of parameter 2 for tone0 and tone1 are quite similar, i.e., in Fig. 5. It has been noted that some distributions have multi-modals, i.e., in Fig. 8 and 10 (parameters 5 and 7, respectively). All in all, in nearly all of the frequency distribution graphs of five tones are distinguished from each other empirically.

From the Table 1, it represents the mean and standard deviation values for all seven parameters in comparison. The parameters of different tones have different levels of mean values and standard deviation values. To distinguish one tone from the others, it is needed to use the derived parameters compositely.

From the experimental results, it has been noted that the selected parameters is needed to further apply with other speech technology. For examples, the parameters are expected to be applied in the tree-based context clustering in Thai speech synthesis (Chomphan and Kobayashi, 2007; Hassini *et al*., 2009; Jenq *et al*., 2009; Teymourzadeh *et al*., 2010) to categorize the speech units into tone groups. The data sharing in each of the speech unit clusters can consequently improve the efficiency of the overall synthesis system.

CONCLUSION

This study proposes an analysis of tone-geometrical model parameters for five Thai tones. The tone-geometrical model has been applied to model a syllable-length portion of the F0 contour. The middle, low, falling, high and rising tones been studied. Seven selected parameters from the tone-geometrical model are extracted. The results show that nearly most of the selected parameters can distinguish five tones explicitly. From this finding, the selected parameters are expected to apply in the speech synthesis systems in the future.

ACKNOWLEDGEMENT

The researcher is grateful to NECTEC for providing the TSynC-1 speech database.

REFERENCES

Alshamasin, M.S., 2009. Optimization of the performance of single-phase capacitor-run induction motor. Am. J. Applied Sci., 6: 745-751. DOI: 10.3844/ajassp.2009.745.751

Chomphan, S. and T. Kobayashi, 2007. Implementation and evaluation of an HMM-based Thai speech synthesis system. Proceeding of the 8th Annual Conference of the International Speech Communication Association, (ISCA'07), Antwerp, Belgium, pp: 2849-2852.

Chomphan, S. and T. Kobayashi, 2008. Tone correctness improvement in speaker dependent HMM-based Thai speech synthesis. Speech Commun., 50: 392-404. DOI: 10.1016/j.specom.2007.12.002

Chomphan, S. and T. Kobayashi, 2009. Tone correctness improvement in speaker-independent average-voice-based Thai speech synthesis. Speech Commun., 51: 330-343. DOI: 10.1016/j.specom.2008.10.003

Chomphan, S., 2009. Towards the development of speaker-dependent and speaker-independent hidden Markov model-based Thai speech synthesis. J. Comput. Sci., 5: 905-914. DOI: 10.3844/jcssp.2009.905.914

Chomphan, S., 2010a. Fujisaki's model of fundamental frequency contours for Thai dialects. J. Comput. Sci., 6: 1263-1271. DOI: 10.3844/jcssp.2010.1263.1271

Chomphan, S., 2010b. 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., 2010c. 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., 2010d. Structural modeling of fundamental frequency contour for Thai expressive speech. J. Comput. Sci., 6: 330-335. DOI: 10.3844/jcssp.2010.330.335

Chomphan, S., 2010e. Tone question of tree based context clustering for hidden Markov model based Thai speech synthesis. J. Comput. Sci., 6: 1468-1472. DOI: 10.3844/jcssp.2010.1468.1472

Fujisaki, H. and H. Sudo, 1971. A model for the generation of fundamental frequency contours of Japanese word accent. J. Acoust. Soc. Japan, 57: 445-452.

Fujisaki, H. and S. Ohno, 1998. The use of generative model of F0 contours for multilingual speech synthesis. Proceedings of the International Conference on Spoken Language Processing, (ICSLP'98), Sydney, Australia, pp: 714-717.

Fujisaki, H., K. Hirose, P. Halle and H. Lei, 1990. Analysis and modeling of tonal features in polysyllabic words and sentences of the standard Chinese. Proceeding of the International Conference on Spoken Language Processing, pp: 841-844.

Hassini, A., F. Benabdelouahed, N. Benabadji and A.H. Belbachir, 2009. Active fire monitoring with level 1.5 MSG satellite images. Am. J. Applied Sci., 6: 157-166. DOI: 10.3844/ajassp.2009.157.166

Hiroya, F. and O. Sumio, 2002. A preliminary study on the modeling of fundamental frequency contours of Thai utterances. Proceedings of the 6th International Conference on Signal Processing, Aug. 6-30, Beijing, China, pp: 516-519. DOI: 10.1109/ICOSP.2002.1181106

Jenq, F.L., K.K. Chong and L.Q. Zeng, 2009. Overcoupled response improvement with miniaturizing of rectangular dual-mode filter by slow-wave modified resonator. J. Applied Sci., 9: 2841-2846.

Li, Y., T. Lee and Y. Qian, 2004. Analysis and modeling of F0 contours for Cantonese text-to-speech. Trans. Asian Language Inform. Process., 3: 169-180. DOI: 10.1145/1037811.1037813

Ni, J. and K. Hirose, 2006. Quantitative and structural modeling of voice fundamental frequency contours of speech in Mandarin. Speech Commun., 48: 989-1008. DOI: 10.1016/j.specom.2006.01.002

Saito, T. and M. Sakamoto, 2002. Applying a hybrid intonation model to a seamless speech synthesizer. Proceeding of the International Conference on Spoken Language Processing, Sept. 16-20, Colorado, USA., pp: 165-168.

Seresangtakul, P. and T. Takara, 2002. Analysis of pitch contour of Thai tone using Fujisaki's model. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Apr. 27-30, Minneapolis, MN, USA., pp: 1-1. DOI: 10.1109/ICASSP.2002.1005787

Seresangtakul, P. and T. Takara, 2003. A generative model of fundamental frequency contours for polysyllabic words of Thai tones. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Apr. 6-10, Okinawa, Japan, pp: 452-455. DOI: 10.1109/ICASSP.2003.1198815

Tao, J., J. Yu and W. Zhang, 2006. Internal dependence based F0 model for mandarin tts system. Proceedings of theTC-STAR Workshop on Speech-to-Speech Translation, June 19-21, Barcelona, Spain, pp: 171-174.

Teymourzadeh R., Y.S. Algnabi, M. Othman, M.S. Islam and J.M.V. Hong, 2010. VLSI implementation of novel class of high speed pipelined digital signal processing filter for wireless receivers. Am. J. Eng. Applied Sci., 3: 663-669. DOI: 10.3844/ajeassp.2010.663.669

Zainal, M.R.M., S.A. Samad, A. Hussain and C.H. Azhari, 2009. Pitch and timbre determination of the angklung. Am. J. Applied Sci., 6: 24-29. DOI: 10.3844/ajassp.2009.24.29