American Journal of Applied Sciences, 2012, 9 (12), 1990-2003
ISSN: 1546-9239
©2012 Science Publication
doi:10.3844/ajassp.2012.1990.2003 Published Online 9 (12) 2012 (http://www.thescipub.com/ajas.toc)

MODELING OF FUNDAMENTAL FREQUENCY CONTOURS FOR THAI DIALECTS WITH LARGE SPEECH DATABASE

[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

Received 2012-03-04, Revised 2012-08-30; Accepted 2012-12-20

ABSTRACT

In four core regions of Thailand, there are four main dialects including central, north, northeast and south dialects. The prosody is significantly unique for each dialect. One important factor determining the prosody is the fundamental frequency. As a result, modeling of Fundamental frequency (F0) contour is very important for the natural speech processing. Even though there are many modeling techniques for modeling the F0 contour. In this study, the Fujisaki's model has been selected because of its achievement in modeling of various Thai speech units. This study proposes an analysis of model parameters of Thai speech prosody for four regional dialects and two genders. Seven derived parameters from the Fujisaki's model are as follows. The first parameter is baseline frequency which is the lowest level of F0 contour. The second and third parameters are the numbers of phrase commands and tone commands which reflect the frequencies of surges of the utterance in global and local levels, respectively. The fourth and fifth parameters are phrase command and tone command durations which reflect the speed of speaking and the length of a syllable, respectively. The sixth and seventh parameters are amplitudes of phrase command and tone command which reflect the energy of the global speech and the energy of local syllable. In the experimental results, the large speech material of each regional dialect includes 50 samples of 50 sentences with male and female speech. It can be obviously seen that most of the proposed parameters can distinguish four kinds of regional dialects explicitly. The results reveal that the proposed parameters of Fujisaki's model can distinguish the regional dialects explicitly.

Keywords: Fundamental Frequency (F0), Regional Dialects Explicitly, Compact Speech Database, North East Dialect, Conventional Parameters, Command Duration

1. INTRODUCTION

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; Hiroya and Hiroshi, 1971; Dat *et al*., 2006). 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). To study how efficient the Fujisaki's model perform in each of Thai dialects (central, north, northeast and south dialects), it has been adapted to the same utterances for all dialects. An analysis of model parameters of Thai speech prosody for four regional dialects and two genders will be performed in the same way as modeling of fundamental frequency for Thai expressive speech conducted in 2010 which is proved to be effective for a limited-domain speech corpus (Chomphan, 2010a). The previous study shows that the derived parameters can distinguish one style of speech from each other.

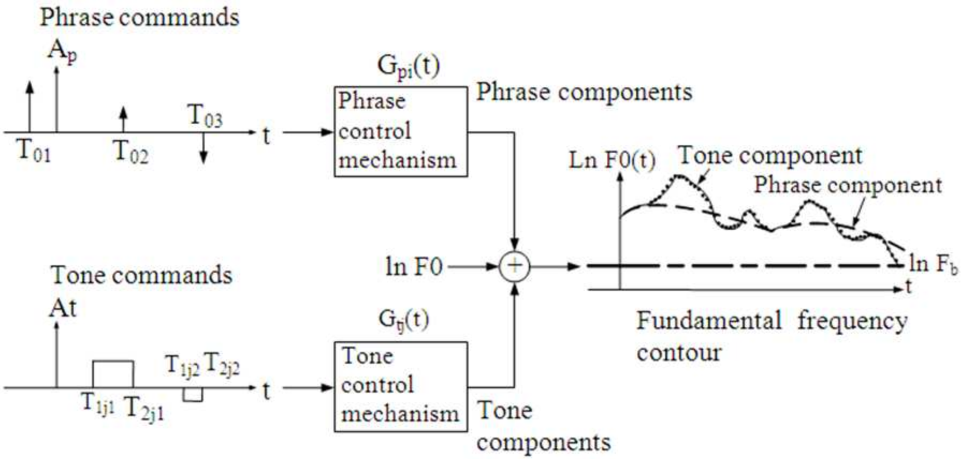


Fig. 1. An extension of Fujisaki's model for the generation of F0 contour

Fujisaki's Modeling of F0 contours for Thai Dialects with a compact speech database has been conducted by Chomphan (2010b). However, the significant differences among dialects cannot be noticed. This study increases the speech material size to 25 times higher than that of the previous study. The study proposes an analysis of F0 modeling of four Thai dialects including standard Thai, Lanna or North dialect, Lao-style or North East dialect and South dialect. The extension of Fujisaki's model which is a preliminary work for the advanced research in speech synthesis and recognition is mainly selected in this study (Seresangtakul and Takara, 2002; 2003).

2. MATERIALS AND METHODS

2.1. Fujisaki's Model

The fundamental frequency contour of an utterance of human speech is treated as a linear superposition of a global phrase and local accent components on a logarithmic scale, as depicted in **Fig. 1** (Hiroya and Hiroshi, 1971).

By using this generative model, the parameters are extracted from our speech database, utterance by utterance. Subsequently, the derived parameters are computed are analyzed (Chomphan and Kobayashi, 2008; 2009; Seresangtakul and Takara, 2003).

2.2. Derived Parameters

From the conventional parameters, we calculated seven derived parameters which reflect the geometrical appearance of the F0 contour of an utterance as follows:

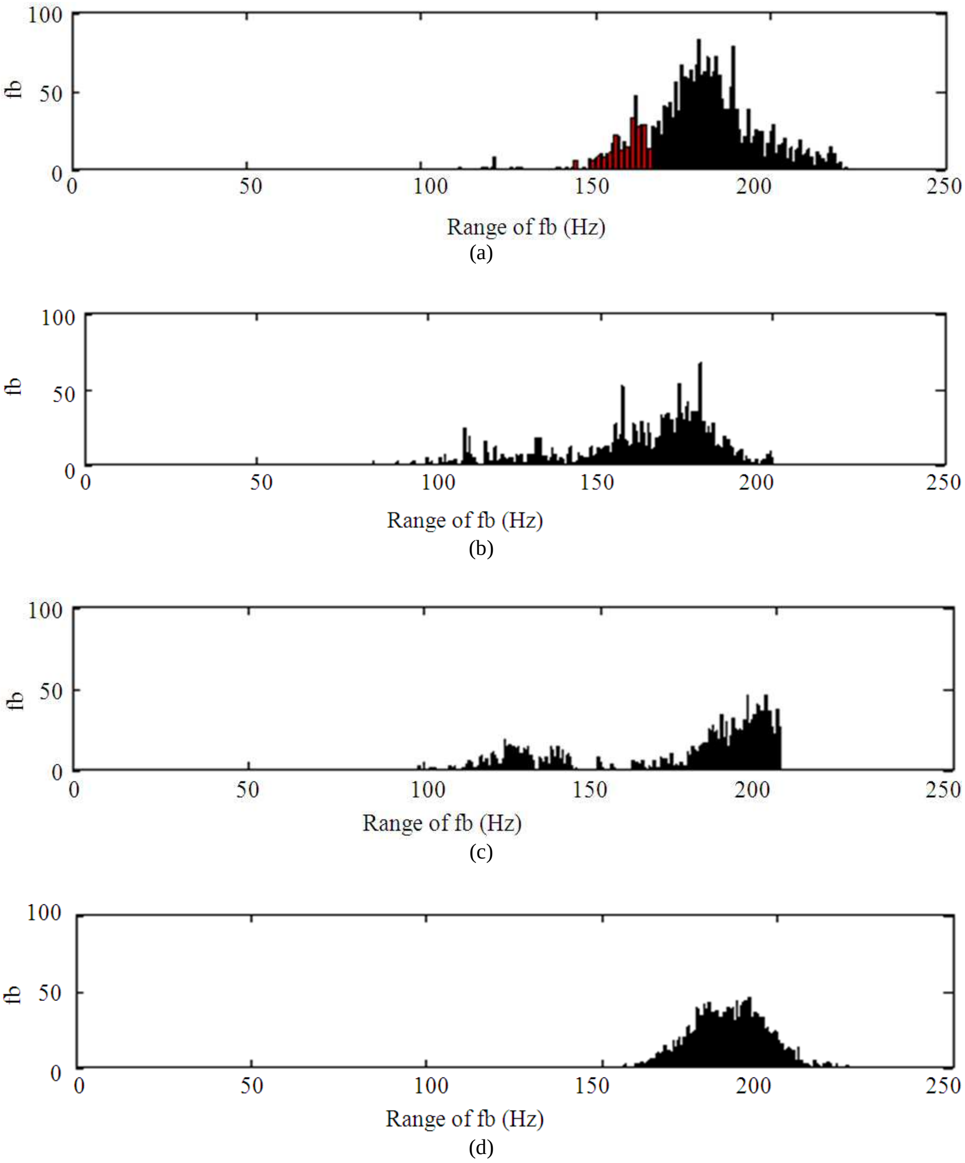
- Baseline frequency
- Number of phrase commands
- Number of tone commands
- Phrase command duration
- Tone command duration
- Amplitude of phrase command
- Amplitude of tone command

All of these derived parameters have been extracted for four regional Thai dialects including standard Thai, Lanna or North dialect, Lao-style or North East dialect and South dialect.

3. RESULTS

In our large speech database, we use fifty sentences in Thai for male and female genders. The sentences have been recorded in four Thai dialects of standard Thai (Center-dialect), Lanna Thai dialect (North-dialect), Lao-style Thai dialect (Northeast-dialect) and South Thai dialect (South-dialect). Each dialect contains two thousands and five hundred utterances of samples. Therefore we have ten thousands utterances of samples for each gender. The parameter extraction tools as used in (Mixdorff and Fujisaki, 1997; Chomphan and Kobayashi, 2007a; 2007b) are exploited in this study.

In each derived parameter, we analyzed the frequency distribution over its range and then the distributions of four Thai dialects including Center dialect, North dialect, Northeast dialect and South dialect are plot in a graph to show the differences and similarities among those dialects. The first seven graphs are of female speech (**Fig. 2-8**) for the baseline frequency, number of phrase commands, number of tone commands, phrase command duration, tone command duration, amplitude of phrase command and amplitude of tone command, respectively. The second seven graphs are of male speech (**Fig. 9-15**) for the baseline frequency, number of phrase commands, number of tone commands, phrase command duration, tone command duration, amplitude of phrase command and amplitude of tone command, respectively. The abbreviations are defined and used in most figures; frame num, fb, PC num, AC num, PC delta t, AC delta t, PC amplitude and AC amplitude, mean number of frames, baseline frequency, number of phrase commands, number of tone commands, phrase command duration, tone command duration, amplitude of phrase command and amplitude of tone command, respectively.



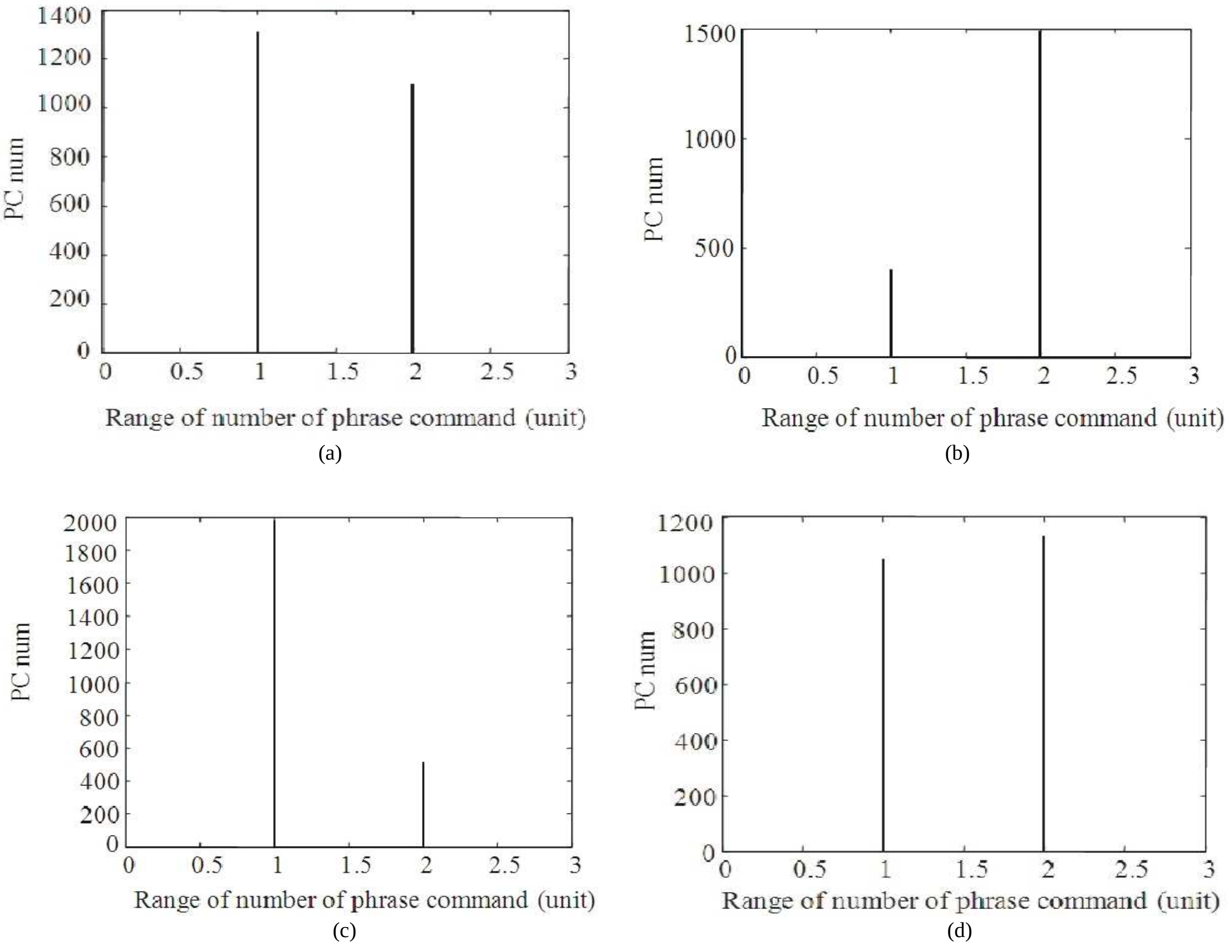
(a)

(b)

(c)

(d)

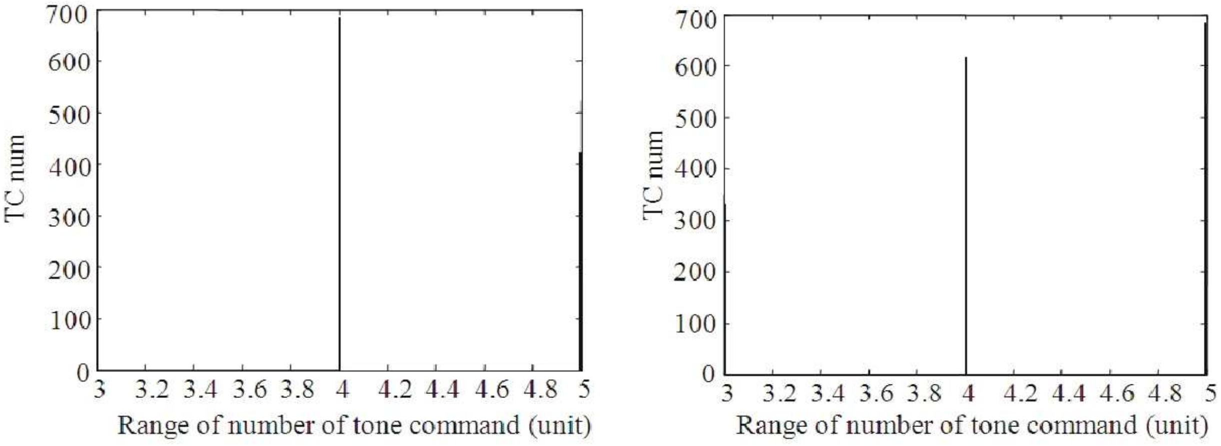
Fig. 2. Comparison of baseline frequency parameter distributions of female for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



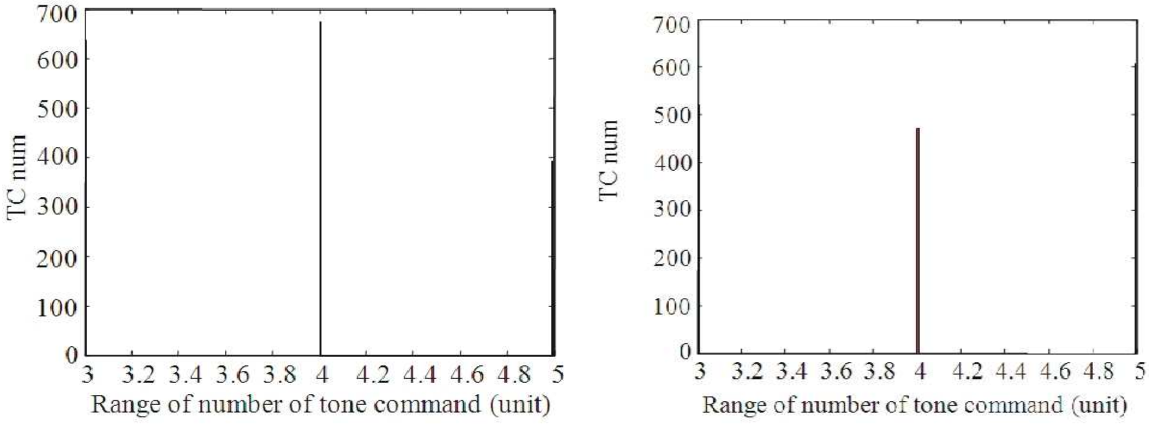
(a) (b)

(c) (d)

Fig. 3. Comparison of number of phrase commands parameter distributions of female for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



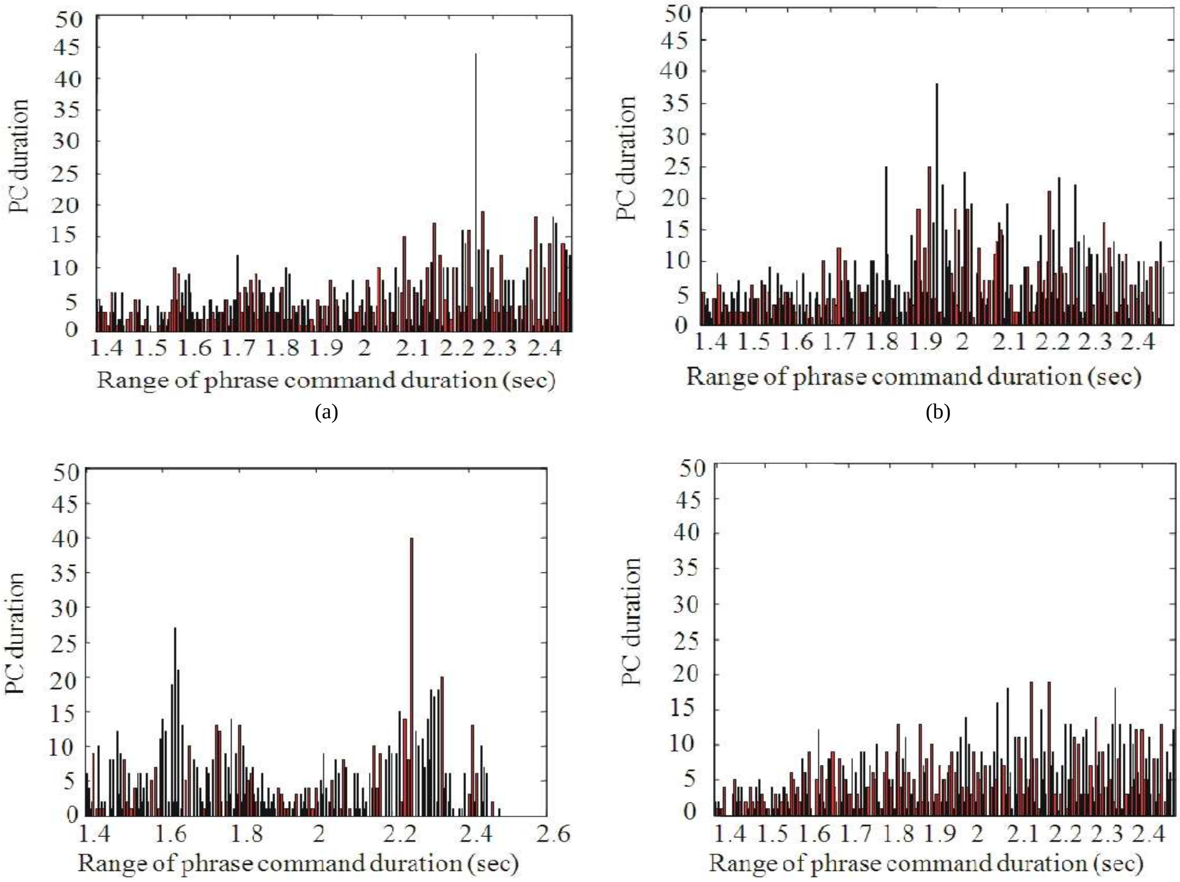
(a) (b)



(c)

(d)

Fig. 4. Comparison of number of tone commands parameter distributions of female for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



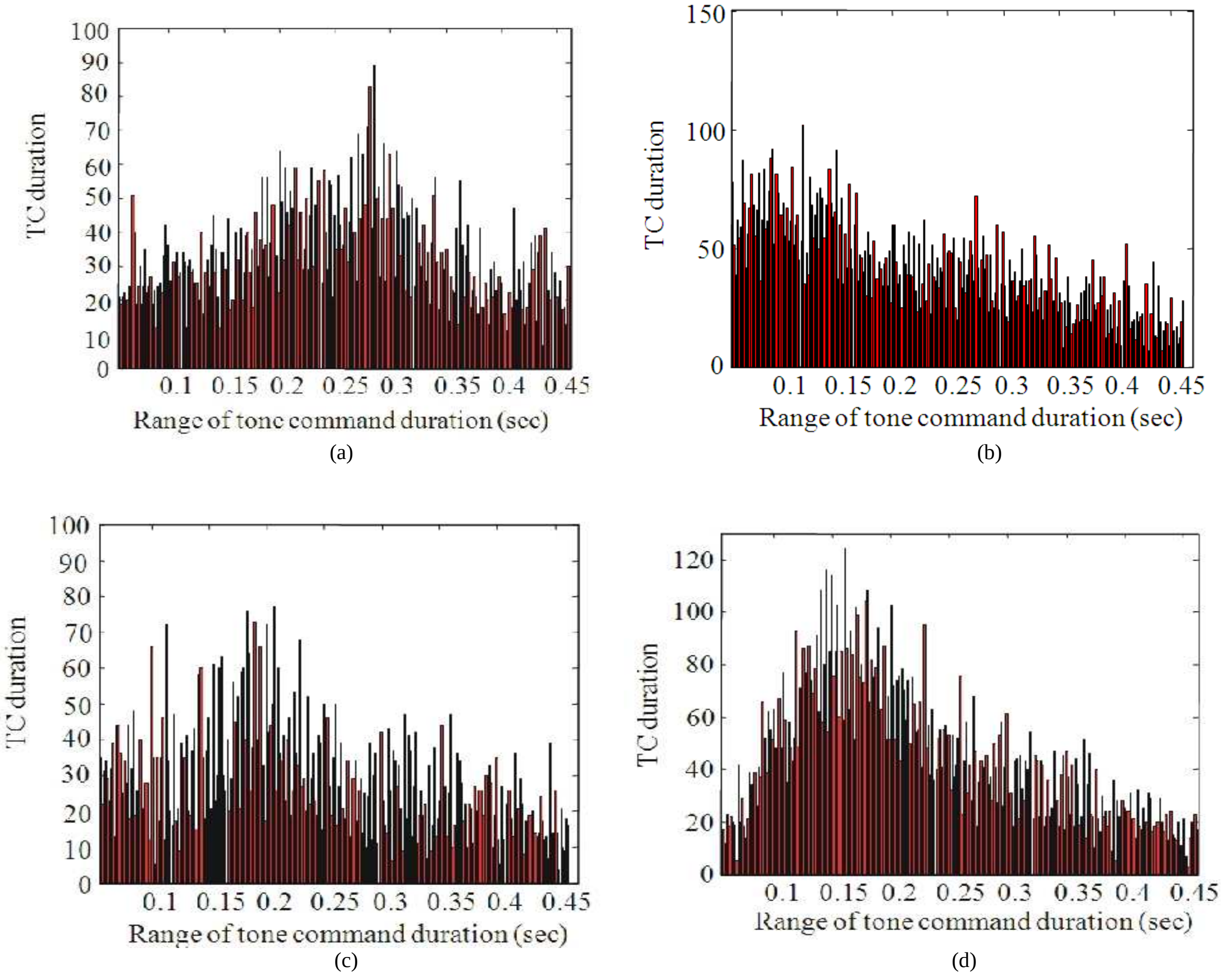
(a)

(b)

(c)

(d)

Fig. 5. Comparison of phrase command duration parameter distributions of female for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



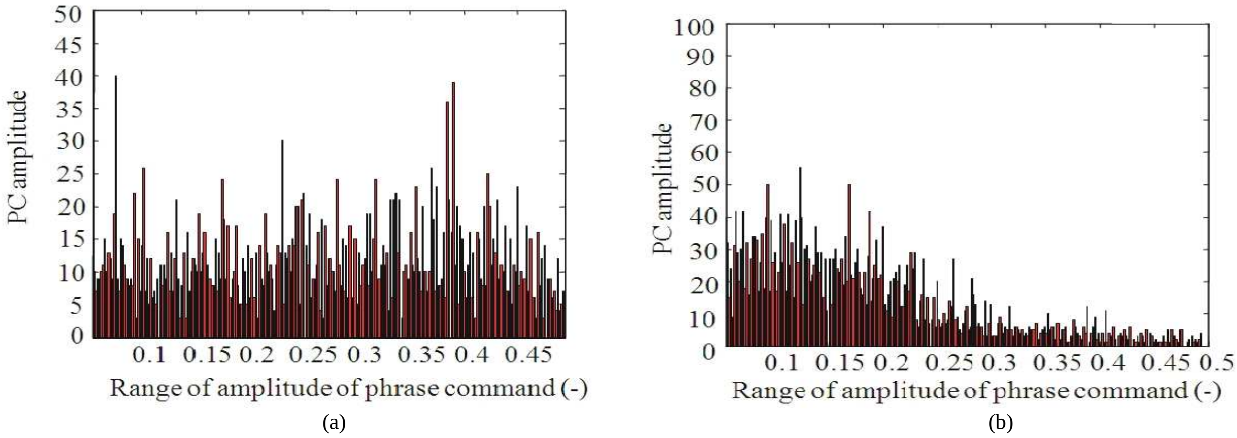
(a)

(b)

(c)

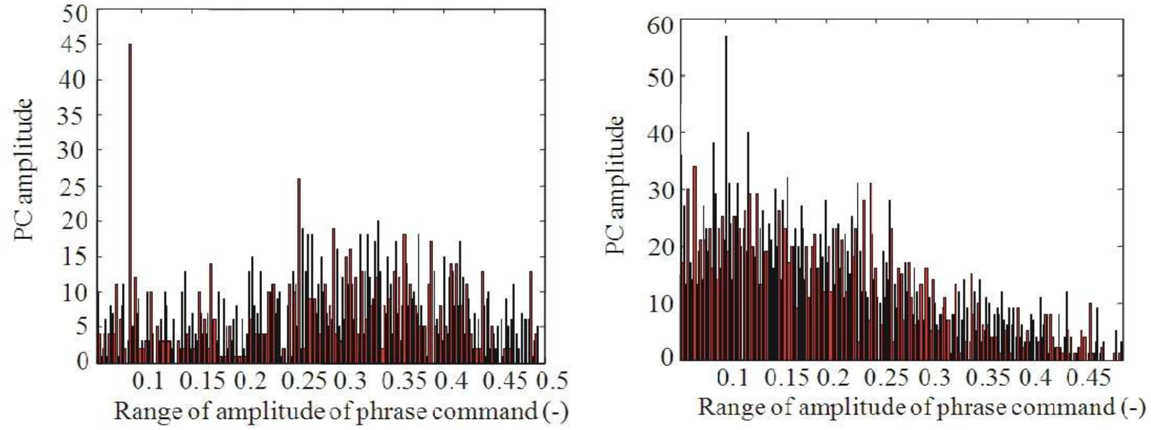
(d)

Fig. 6. Comparison of tone command duration parameter distributions of female for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



(a)

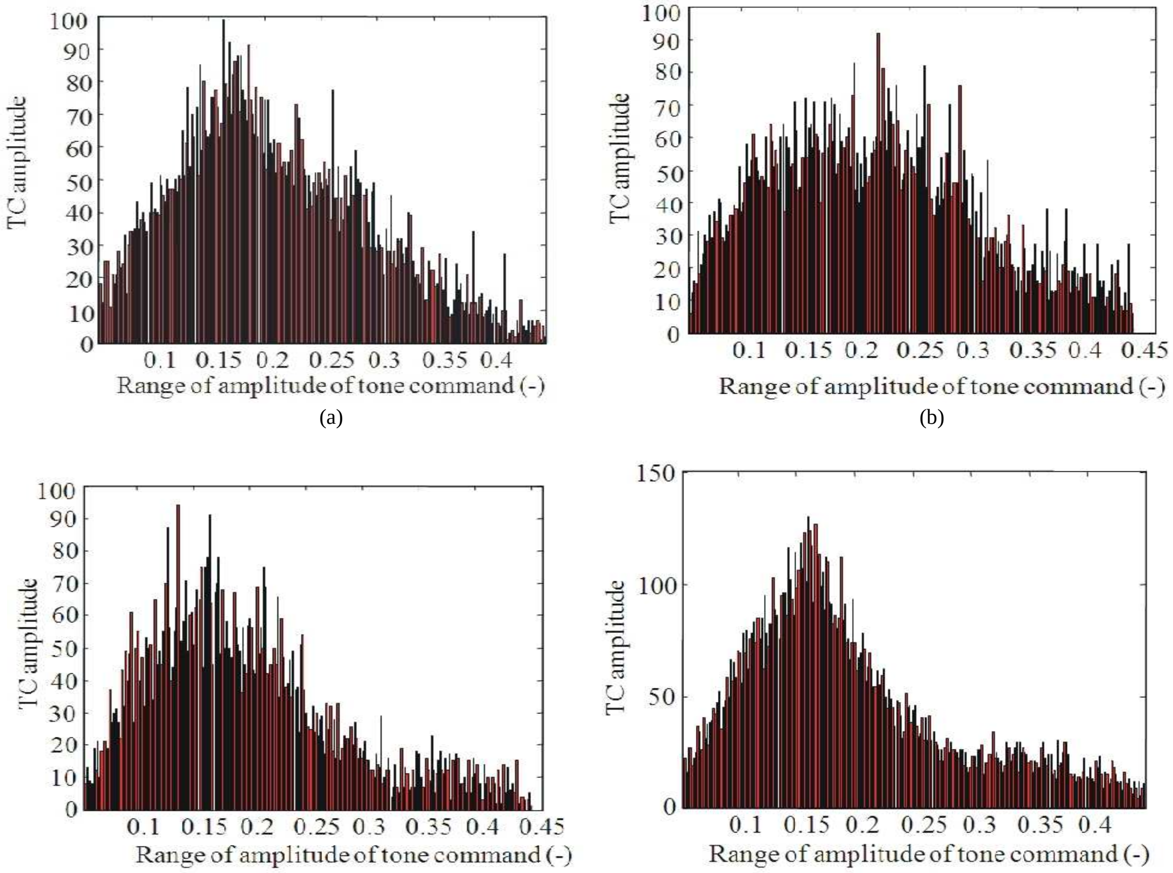
(b)



(c)

(d)

Fig. 7. Comparison of phrase command amplitude parameter distributions of female for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



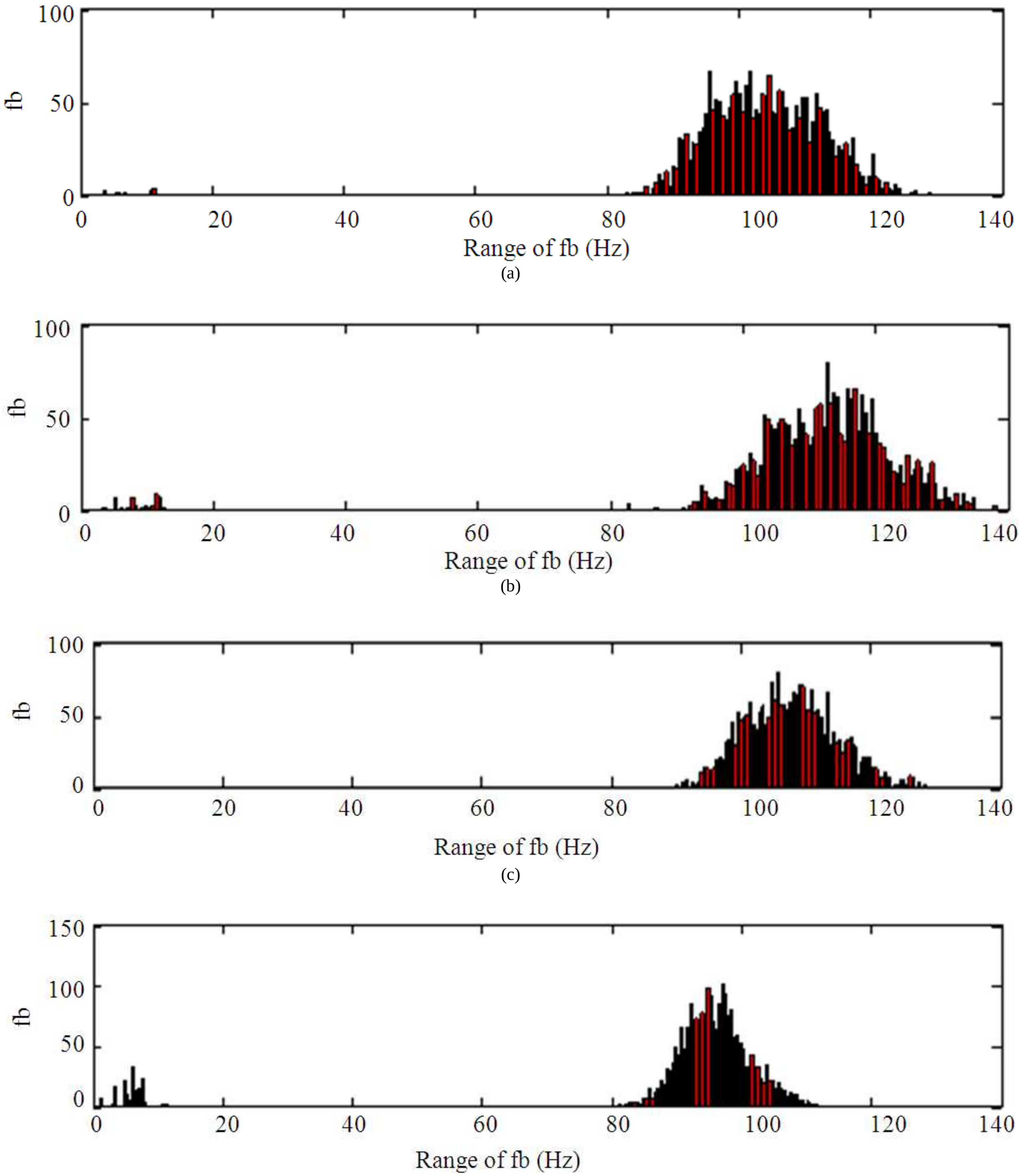
(a)

(b)

(c)

(d)

Fig. 8. Comparison of tone command amplitude parameter distributions of female for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



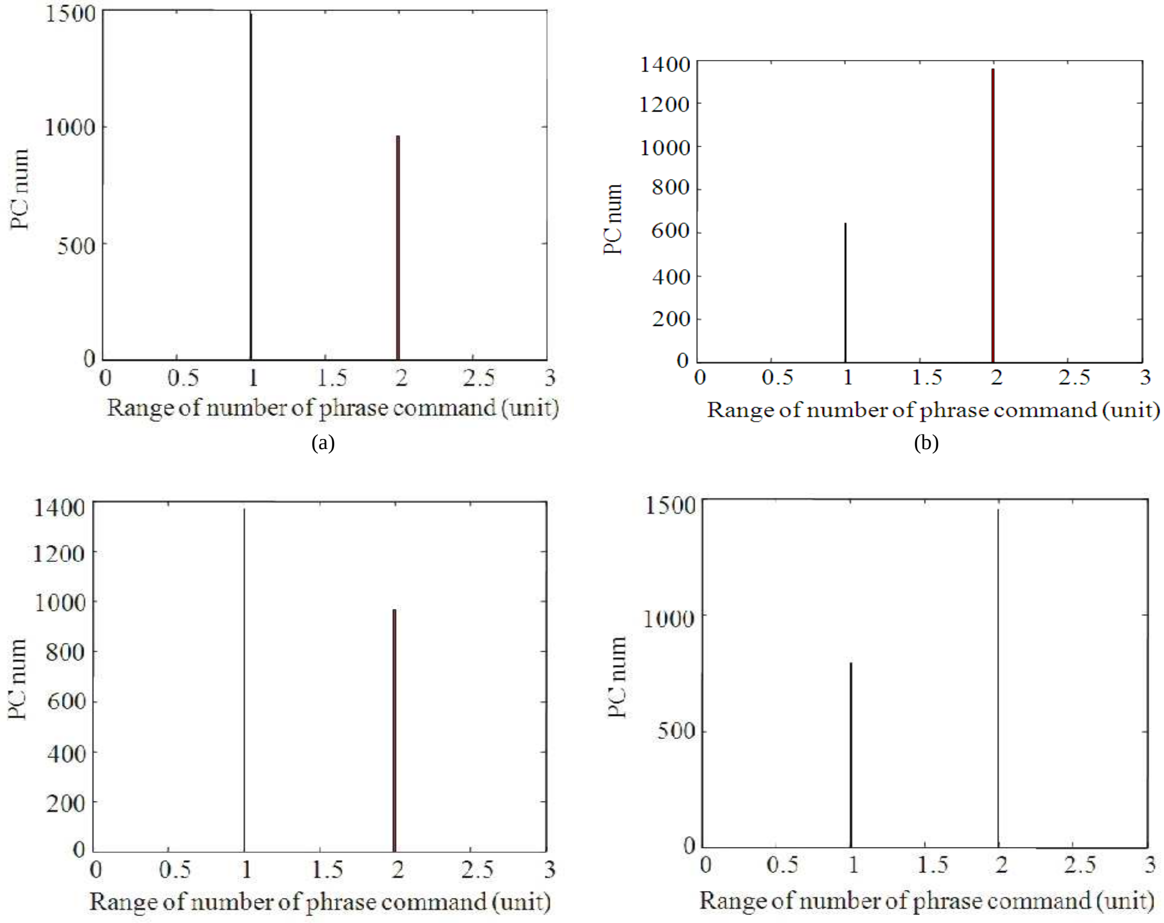
(a)

(b)

(c)

(d)

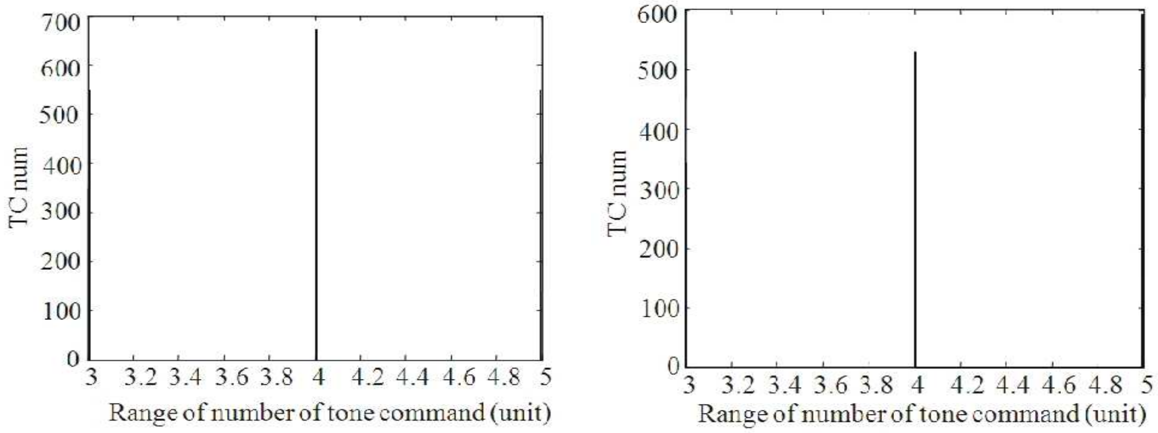
Fig. 9. Comparison of baseline frequency parameter distributions of male for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



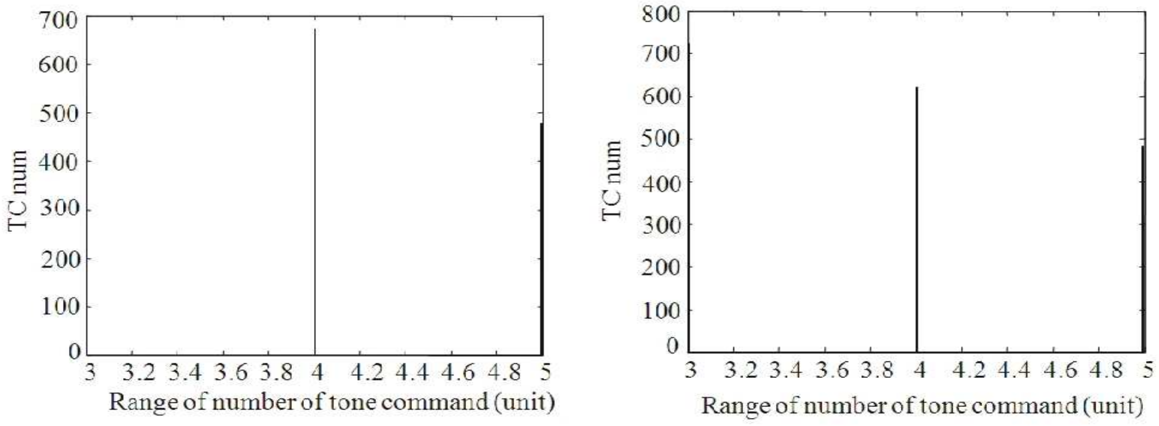
(a) (b)

(c) (d)

Fig. 10. Comparison of number of phrase commands parameter distributions of male for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



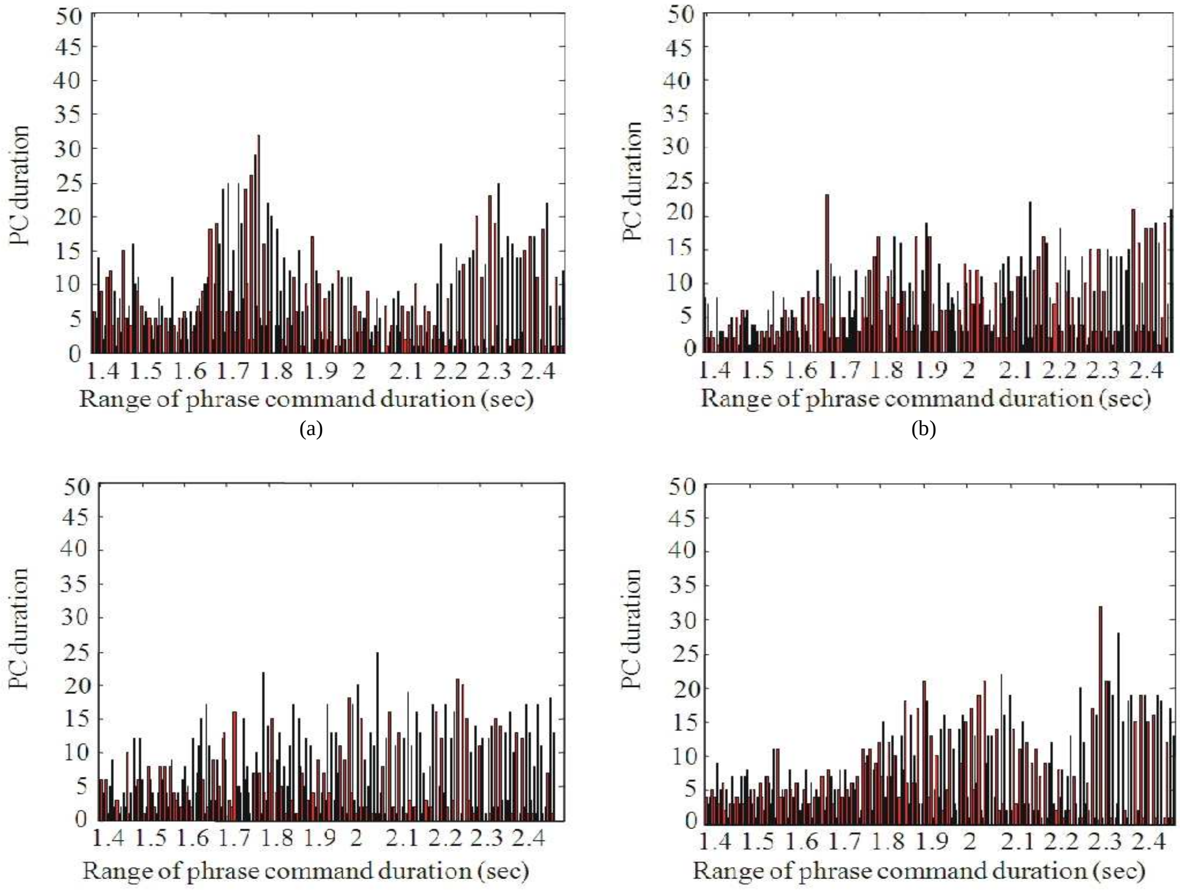
(a) (b)



(c)

(d)

Fig. 11. Comparison of number of tone commands parameter distributions of male for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



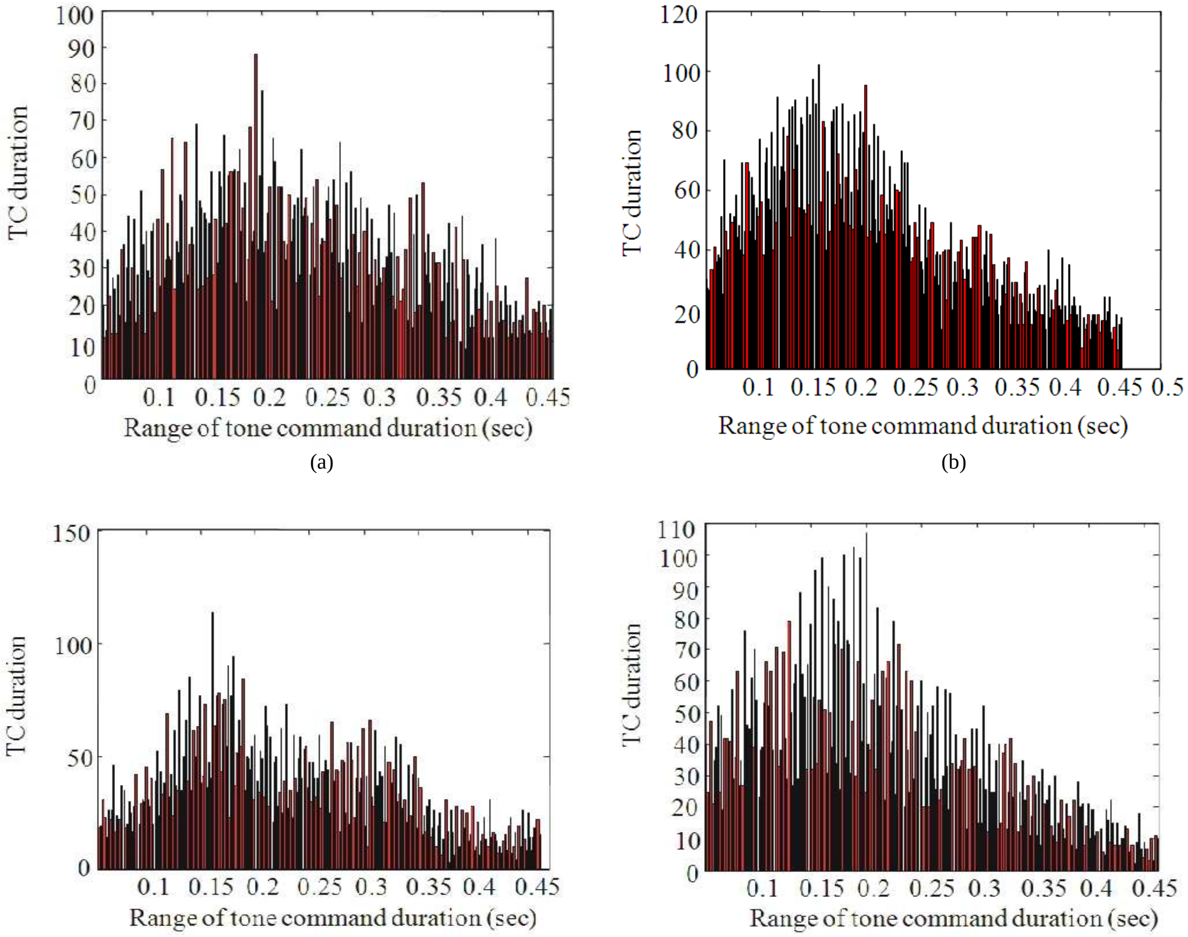
(a)

(b)

(c)

(d)

Fig. 12. Comparison of phrase command duration parameter distributions of male for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



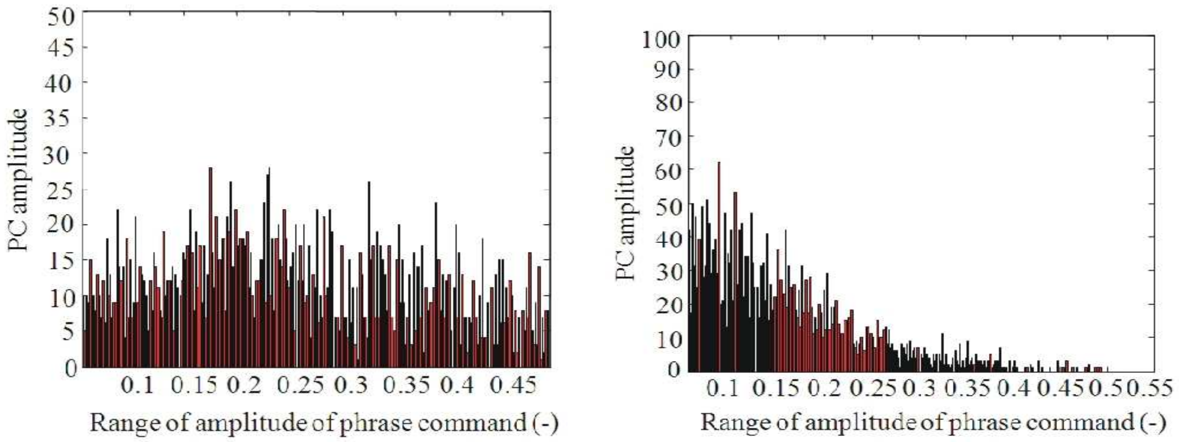
(a)

(b)

(c)

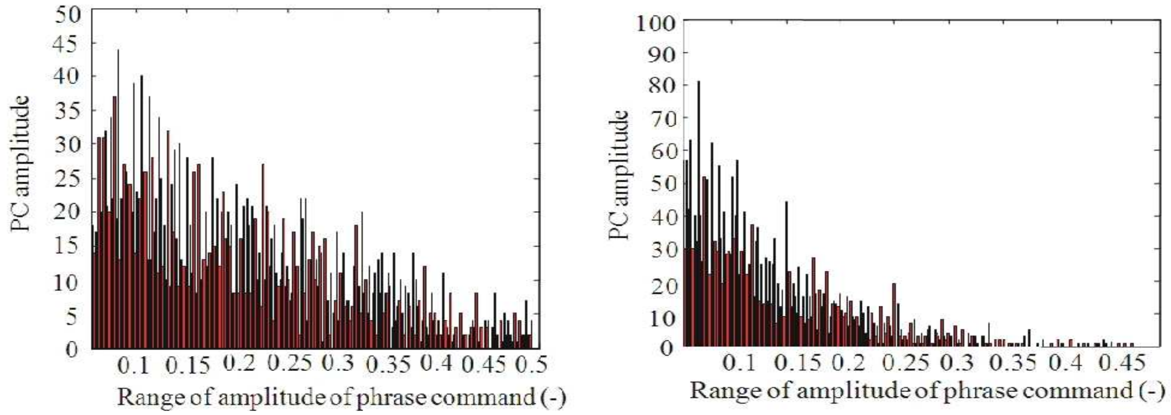
(d)

Fig. 13. Comparison of tone command duration parameter distributions of male for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



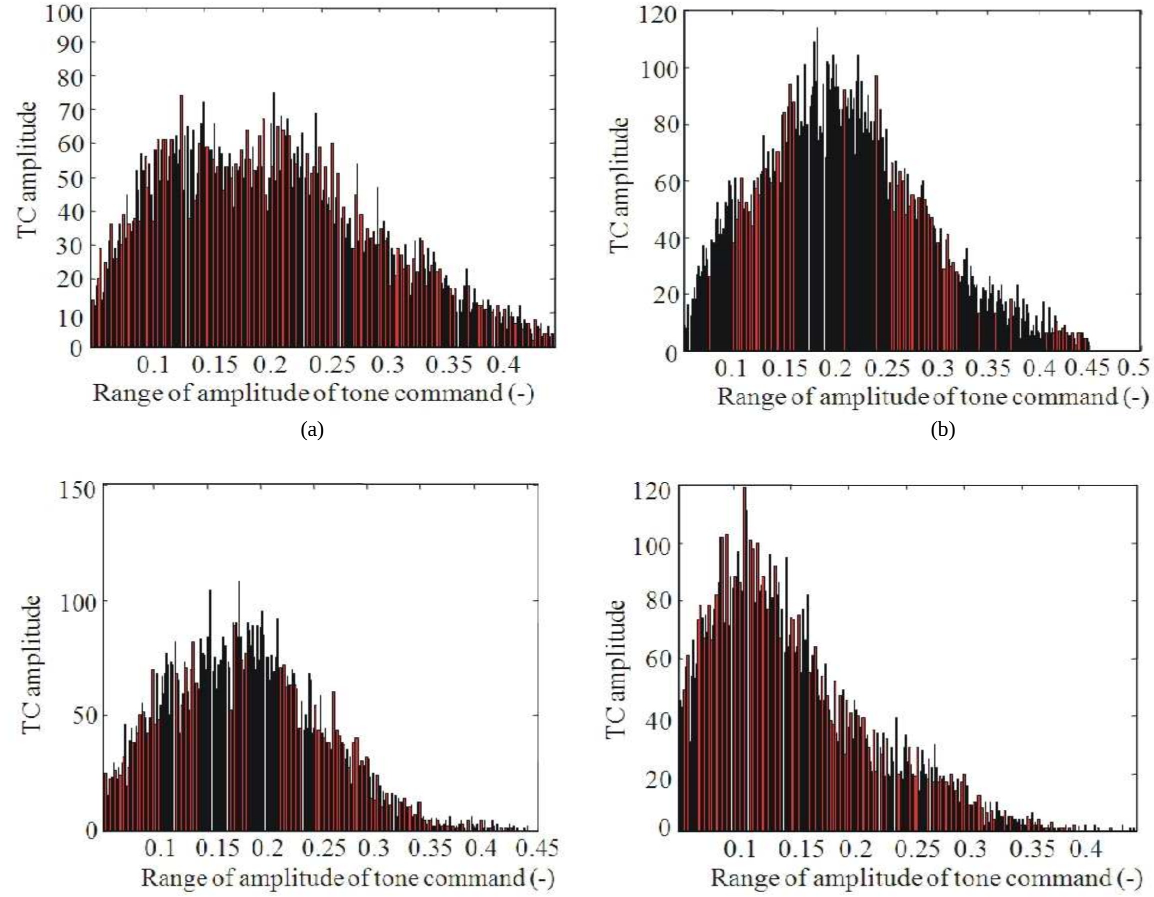
(a)

(b)



(c) (d)

Fig. 14. Comparison of phrase command amplitude parameter distributions of male for four Thai dialects; (a) Center (b) North (c) Northeast (d) South



(a) (b)

(c) (d)

Fig. 15. Comparison of tone command amplitude parameter distributions of male for four Thai dialects; (a) Center (b) North (c) Northeast (d) South

4. DISCUSSION

It can be seen from the frequency distribution graphs of female and male speech in **Fig. 2-15** that most results show that the distributions of four dialects are significantly different. As for the parameter of baseline frequency of female speech in **Fig. 2**, it can be empirically seen that deviation of the south dialect is lowest, while the highest mean value is of the northeast dialect. As for the parameter of number of phrase commands of female speech in **Fig. 3**, it can be noticed that the mode value is at 1 for the center and the northeast dialects, meanwhile the mode value is at 2 for the north and the south dialects. As for the parameter of tone command amplitude of female speech in **Fig. 8**, it can be noticed that the highest mean value is of the northeast dialect, meanwhile the mean values of the other dialects are somewhat similar.

As for the parameter of baseline frequency of male speech in **Fig. 9**, it can be empirically seen that deviation of the south dialect is lowest, while the highest mean value is of the north dialect. As for the parameter of number of phrase commands of male speech in **Fig. 10**, it can be noticed that the mode value is at 1 for the center and the northeast dialects, meanwhile the mode value is at 2 for the north and the south dialects. As for the parameter of tone command amplitude of male speech in **Fig. 15**, it can be noticed that the highest mean value is of the northeast dialect, meanwhile the lowest mean value is of the south dialect.

5. CONCLUSION

In this study, the study of a modeling of F0 contour for Thai dialects with a large speech database is conducted. The Fujisaki's model which is proved to be efficient for several Thai speech units has been chosen in this study. The differences among the model parameters of four Thai dialects have been discussed. The experimental results indicate that most of the proposed parameters can distinguish four kinds of Thai dialects obviously.

6. ACKNOWLEDGEMENT

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

7. REFERENCES

Chomphan, S. and T. Kobayashi, 2007a. Design of tree-based context clustering for an HMM-based Thai speech synthesis system. Proceedings of the 6th ISCA Workshop on Speech Synthesis, Aug. 22-24, Bonn, Germany, pp: 160-165.

Chomphan, S. and T. Kobayashi, 2007b. Implementation and evaluation of an HMM-based Thai speech synthesis system. Proceedings 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., 2010a. Analytical study on fundamental frequency contours of thai expressive speech using Fujisaki's model. J. Comput. Sci., 6: 36-42. DOI: 10.3844/jcssp.2010.36.42

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

Tran, D. D., E. Castelli, J.F. Serignat, V.L. Trinh, X.H. Le, 2006. Linear F0 contour model for Vietnamese tones and Vietnamese syllable synthesis with TD-PSOLA. Proceedings of the 2nd International Symposium on Tonal Aspects of Languages, (TAL' 06).

Fujisaki, H. and S. Ohno, 1998. The use of a generative model of F_0 contours for multilingual speech synthesis. Proceedings of the 4th International Conference on Signal Processing, Oct. 12-16, IEEE Xplore Press, Beijing, pp: 714-717. DOI: 10.1109/ICOSP.1998.770311

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. Proceedings of the International Conference on Spoken Language Processing, (SLP' 90), CiteSeerX, USA.

Hiroya, F. and S. Hiroshi, 1971. A model for the generation of fundamental frequency contours of Japanese word accent. J. Acoust. Soc. Japan, 57: 445-452.

Hiroya, F. and O. Sumio, 2002. A preliminary study on the modeling of fundamental frequency contours of Thai utterances. Proceedings of the International Conference on Signal Processing, Aug. 26-30, IEEE Xplore Press, pp: 516-519. DOI: 10.1109/ICOSP.2002.1181106

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

Mixdorff, H. and H. Fujisaki, 1997. Automated quantitative analysis Of F0 contours of utterances from a German ToBI-labeled speech database. Proceedings of the 5th European Conference on Speech Communication and Technology, Sept. 22-25, Rhodes, Greece, pp: 187-190.

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. Proceedings of the 7th 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 IEEE International Conference on Acoustics, Speech and Signal Processing, May 13-17, IEEE Xplore Press, Orlando, FL, USA., pp: 505-508. DOI: 10.1109/ICASSP.2002.5743765

Seresangtakul, P. and T. Takara, 2003. A generative model of fundamental frequency contours for polysyllabic words of Thai tones. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Apr. 6-10, IEEE Xplore Press, Hong Kong, 2003, 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 the TC-STAR Workshop on Speech-to-Speech Translation, Jun. 19-21, Barcelona, Spain, pp: 171-174.