Analytical Study of Fujisaki's Model of Fundamental Frequency Contour for Thai Tones

[1,2]Suphattharachai Chomphan and [3]Chutarat Chompunth
[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
[3]School of Social and Environmental Development, National Institute of Development
Administration, 118 M.3, Serithai Road, Klong-Chan, Bangkapi, Bangkok, 10240, Thailand

Abstract: Problem statement: Tone of a tonal language is an important feature of a prosodic syllable to identify the meanings of that syllable or that part of word. Ii is very crucial to model the feature related to tone of speech to achieve the most naturalness in speech communication. **Approach:** The study presents an approach to analyze the model parameters of Thai tones for two genders. The successive modeling of fundamental frequency, Fujisaki's model is selected. We derive seven parameters; baseline frequency, the numbers of phrase commands and tone commands, phrase command and tone command durations, amplitudes of phrase command and tone command. **Results:** In the experimental results, there are 20 syllables and each syllable includes 50 samples of a tone with male and female speech. Five tones are recorded in the same environment. Thereafter, there are ten thousands samples in the speech corpus. It can be obviously seen that Thai tones are determined by the derived parameters. **Conclusion:** All in all, Thai tones are able to be discriminated by the derived parameter of Fujisaki's model.

Key words: Fujisaki's model, fundamental frequency, pitch, prosodic information, prosody, Fujisaki's, speech analysis, speech processing

INTRODUCTION

The necessity of fundamental frequency modeling with high accuracy is required in the modern speech processing technology. In the past decades, this related work were performed in many research group of speech technology (Tao *et al*., 2006; Ni and Hirose, 2006; Fujisaki and Sudo, 1971; Saito and Sakamoto, 2002; Li *et al*., 2004; Tran *et al*., 2006; Fujisaki *et al*., 1990; Fujisaki and Ohno, 1998). The speech utterance is divided into many level of speech units. Thereafter, they are modeled by using a lot of modeling techniques (Hiroya and Sumio, 2002). In Thai language, the derived model was developed and applied to many levels of speech units. This study proposes an analytical study of fundamental frequency modeling of tonal language of Thai. We specify on the level of tone speech units (Seresangtakul and Takara, 2002; 2003).

MATERIALS AND METHODS

Fujisaki's model: Figure 1 illustrates an extension of Fujisaki's model for the generation of fundamental frequency contour of a speech utterance. The series of fundamental frequency values is determined as a linear superposition of a local accent component and a global phrase component on a logarithmic scale. (Fujisaki and Sudo, 1971; Chomphan and Kobayashi, 2008). The related parameters are extracted from the speech corpus using the Fujisaki's model for all utterances. Subsequently, the derived output parameters are computed, averaged and systematically analyzed (Chomphan and Kobayashi, 2009; Seresangtakul and Takara, 2003).

Output parameters: Seven derived output parameters are computed from the conventional parameters. It is noticed that these derived paramenters mostly reflect the geometrical appearance of the F0 contour of the speech.

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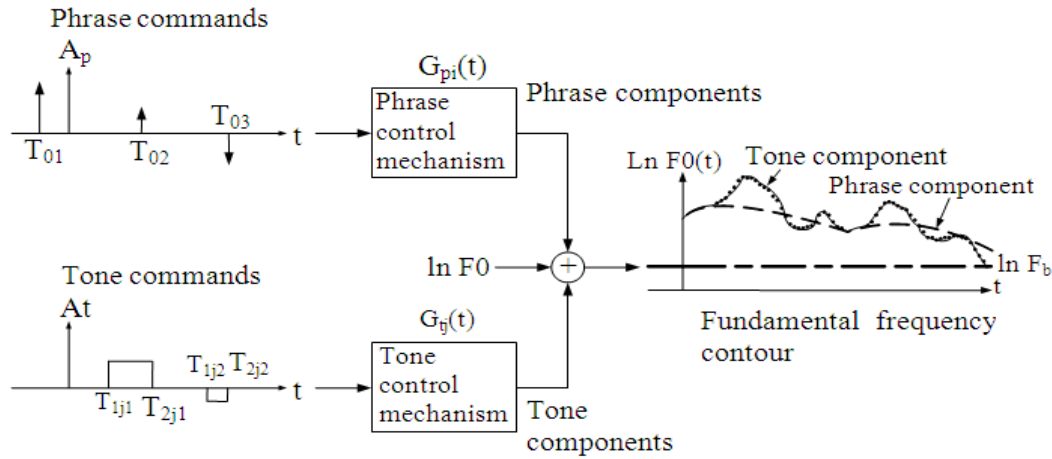
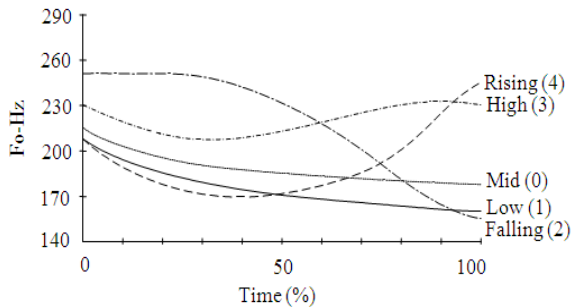Fig. 1: An extension of Fujisaki's model for the generation of F0 contour

Fig. 2: F0 characteristics of 5 tones in Thai

They are 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.
The derived output parameters are mostly extracted for Thai tones (Chomphan, 2011).

To analyzed the distributions of the output parameters statistically, all of output parameters have been extracted for five Thai tones including mid, low, falling, high and rising tones.

Tones in Thai: In tonal language; e.g.,Thai language, tone is a feature embedding within the syllable speech unit. The tone is sometimes interpreted as the accent component (Chomphan and Kobayashi, 2009). However, the number of tones are different for any tonal languauges. In Thai, there are basically 5 tones including mid, low, falling, high and rising tones. The tone numbers are assigned as tone 0, tone 1, tone 2, tone 3 and tone 4, respectively. The trajectories of fundamental frequencies for all tones are different as illustrated in Fig. 2 (Seresangtakul and Takara, 2002; 2003; Chomphan and Kobayashi, 2008).

RESULTS

At first, the Thai speech material is analyzed. Twenty syllables are concerned in this study. Both man speech and woman speech are exploited in the experiment (Mixdorff and Fujisaki, 1997). As for one syllable, five types of tones were covered. Each tone contains 50 samples. Therefore we have ten thousands of samples for both man and woman speech.

In the experiment, the frequency distribution of each output parameter is ploted within its range. Figure 3-7 present the male-speech frequency distributions of all parameters for tone0, tone1, tone2, tone3 and tone4, respectively. Meanwhile, Fig. 8-12 present the female-speech frequency distributions of all parameters for tone0, tone1, tone2, tone3 and tone4, respectively. In the following figures, it has been noted that 'fb', 'pc', 'tc', 'num' denote 'baseline frequency', 'phrase command', 'tone command' and 'number', respectively.

DISCUSSION

We notice from the frequency distribution plots as illustrated in Fig. 3-12. Comparing between five tones, the mean, median and mode of each tone are quite discriminating. At the beginning, the parameter of baseline frequency of male speech in Fig. 3-7, we can empirically noticed that the distribution of tone2 is bi-modal, while the others are uni-modal. Secondly, we can empirically noticed the distributions of parameter of baseline frequency of female speech in Fig. 8-12, the distributions of tone 1 and tone 4 are uni-modal, while the others are not uni-modal.
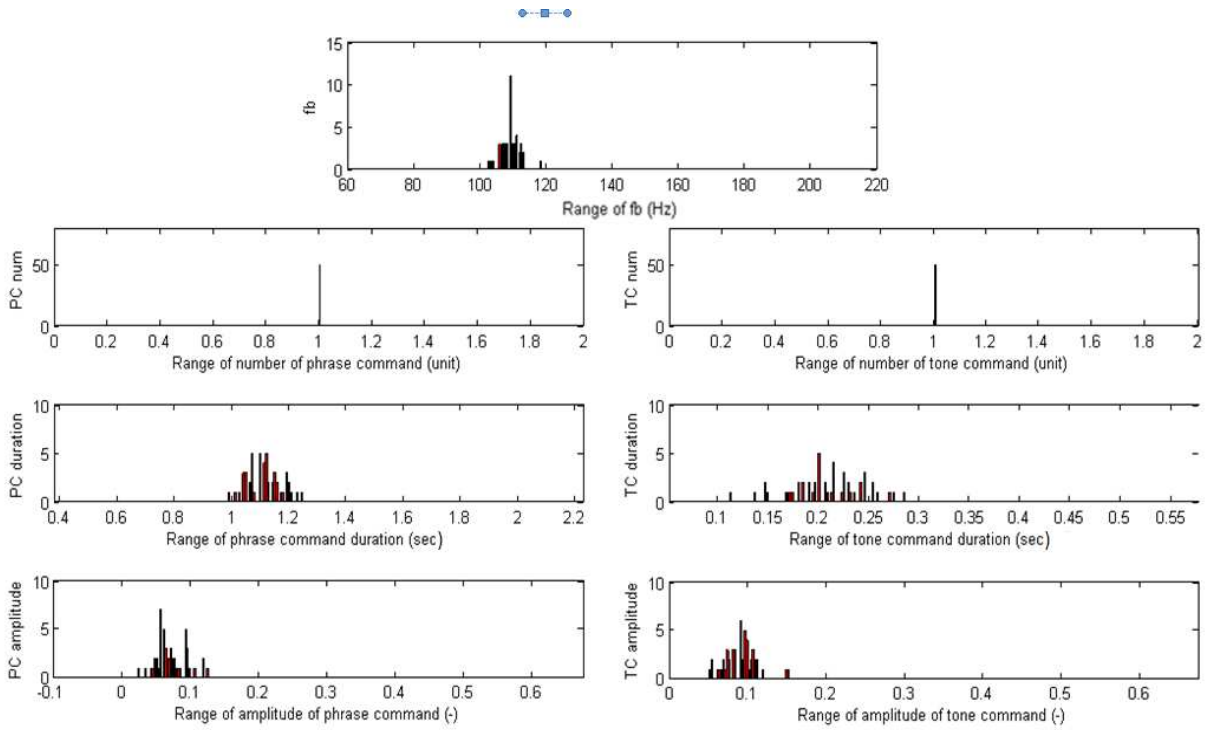
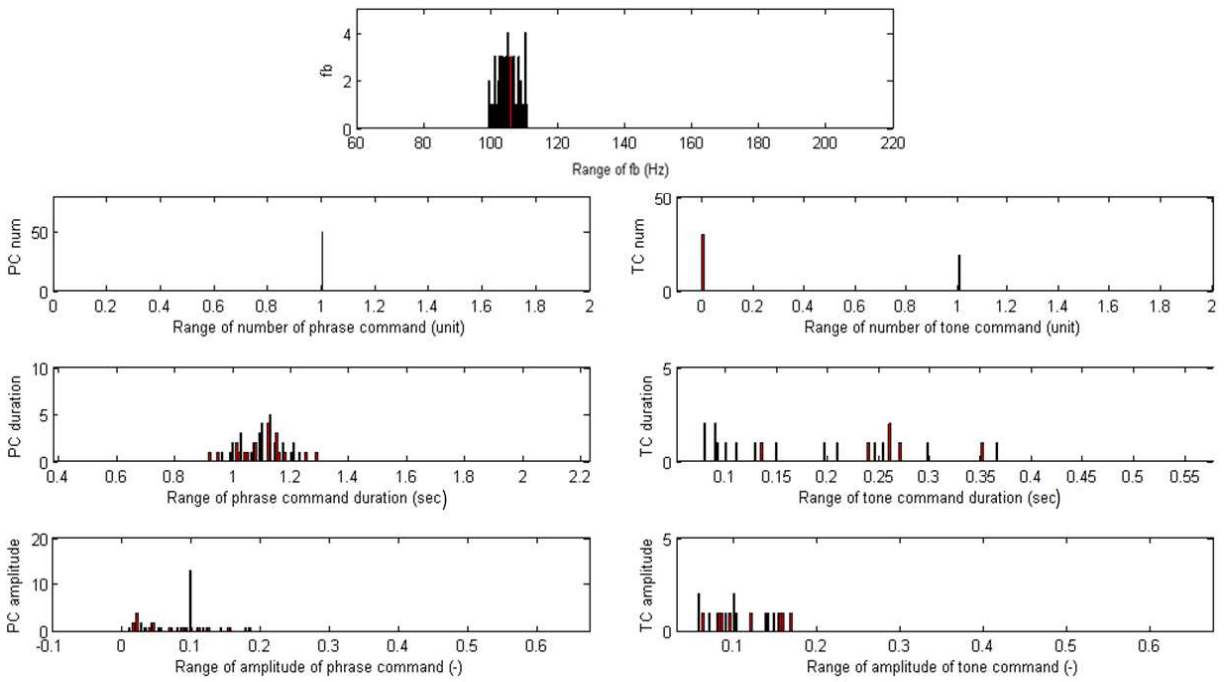Fig. 3: Male-speech frequency distributions of 7 parameters for tone 0

Fig. 4: Male-speech frequency distributions of 7 parameters for tone 1
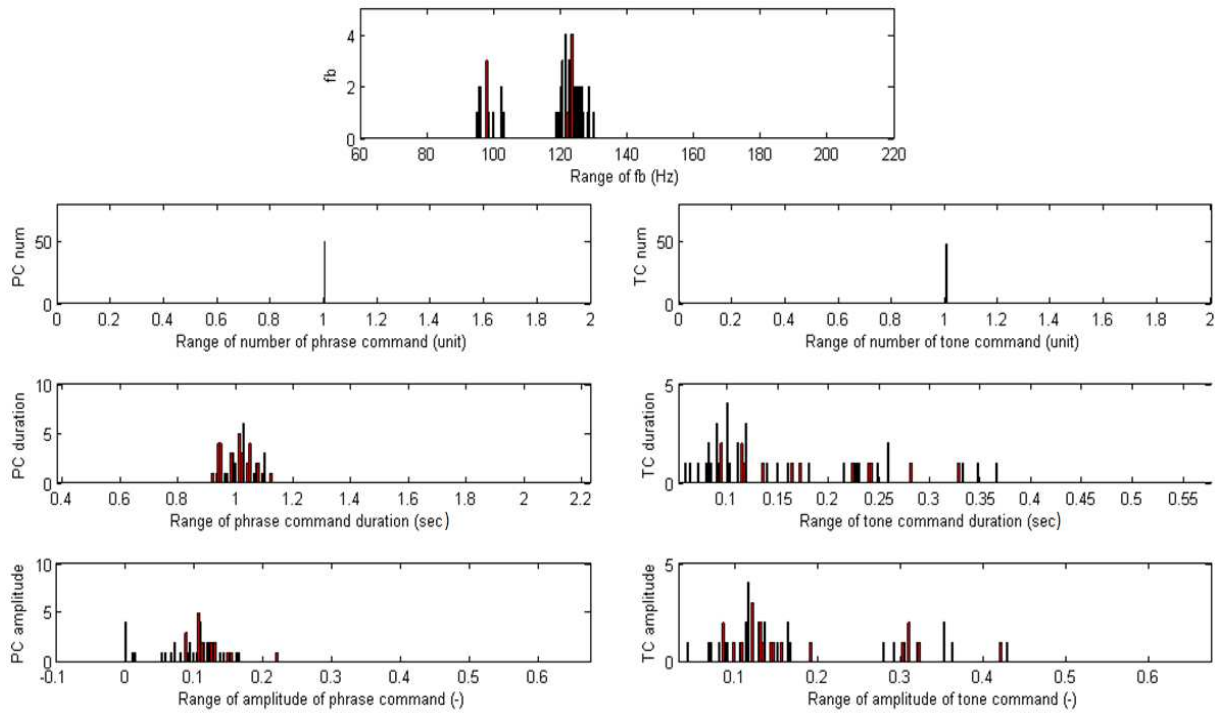
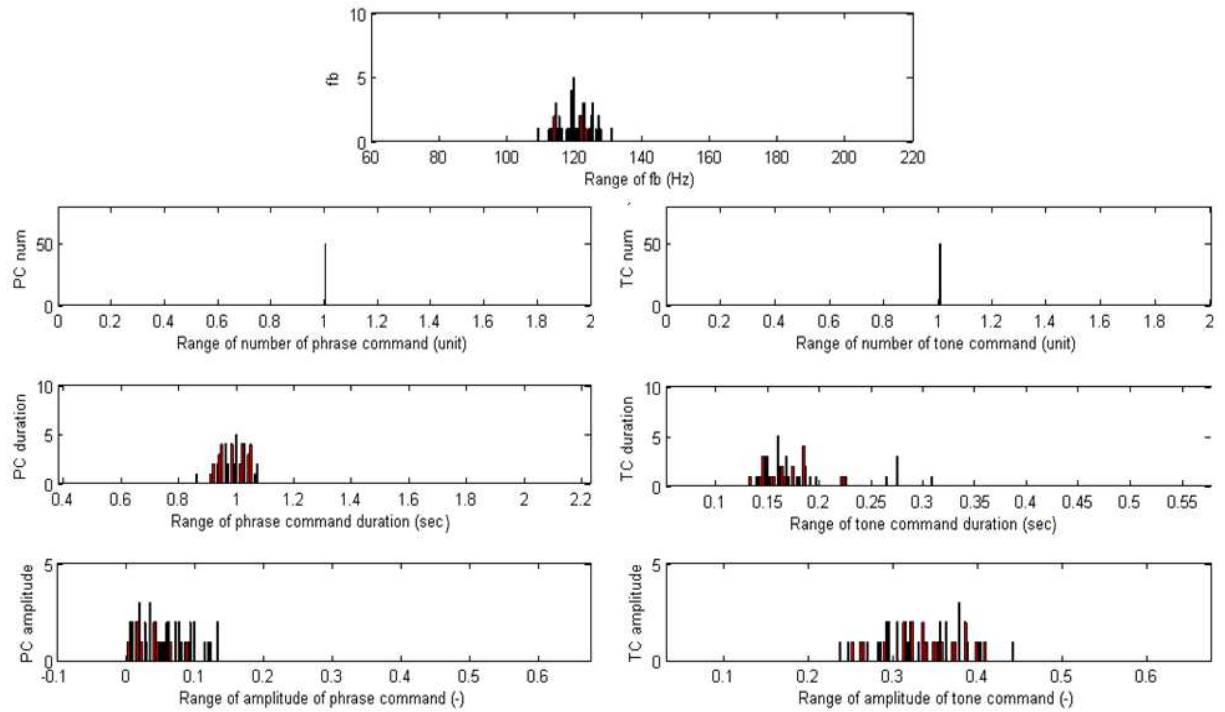Fig. 5: Male-speech frequency distributions of 7 parameters for tone 2

Fig. 6: Male-speech frequency distributions of 7 parameters for tone 3
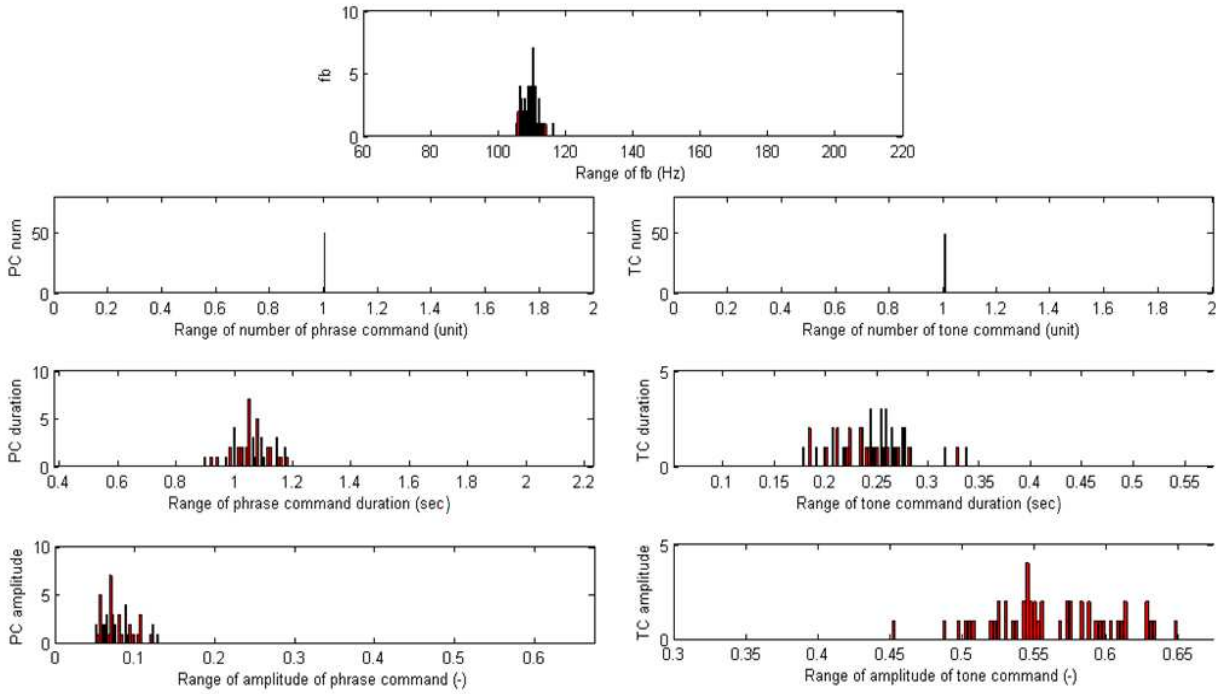
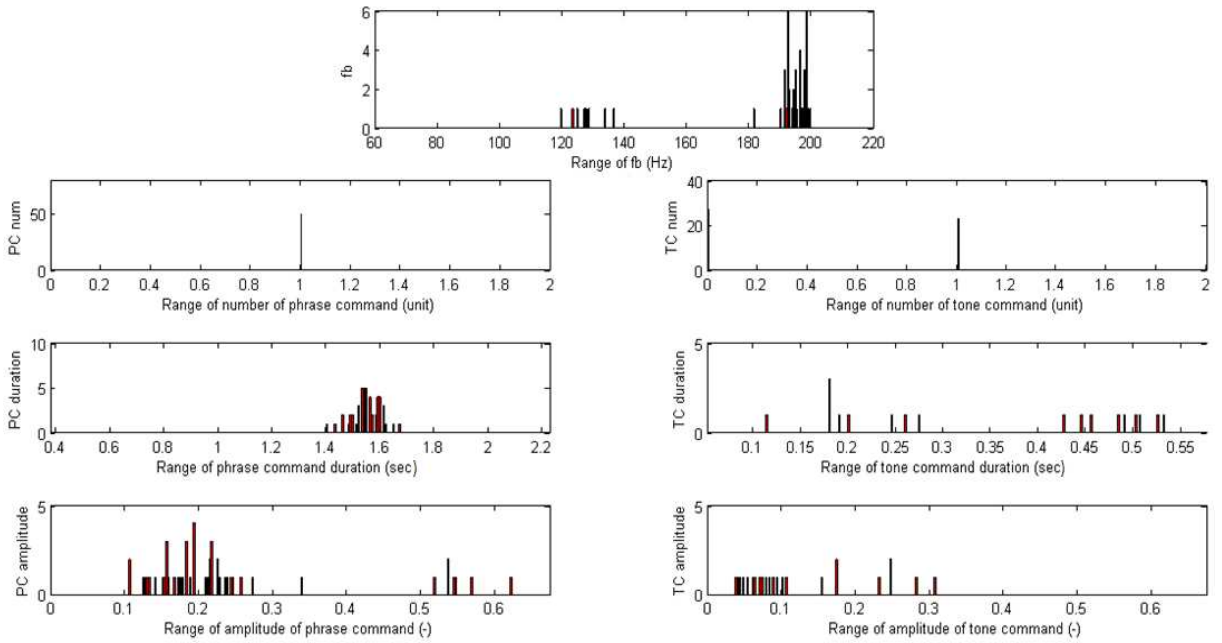Fig. 7: Male-speech frequency distributions of 7 parameters for tone 4

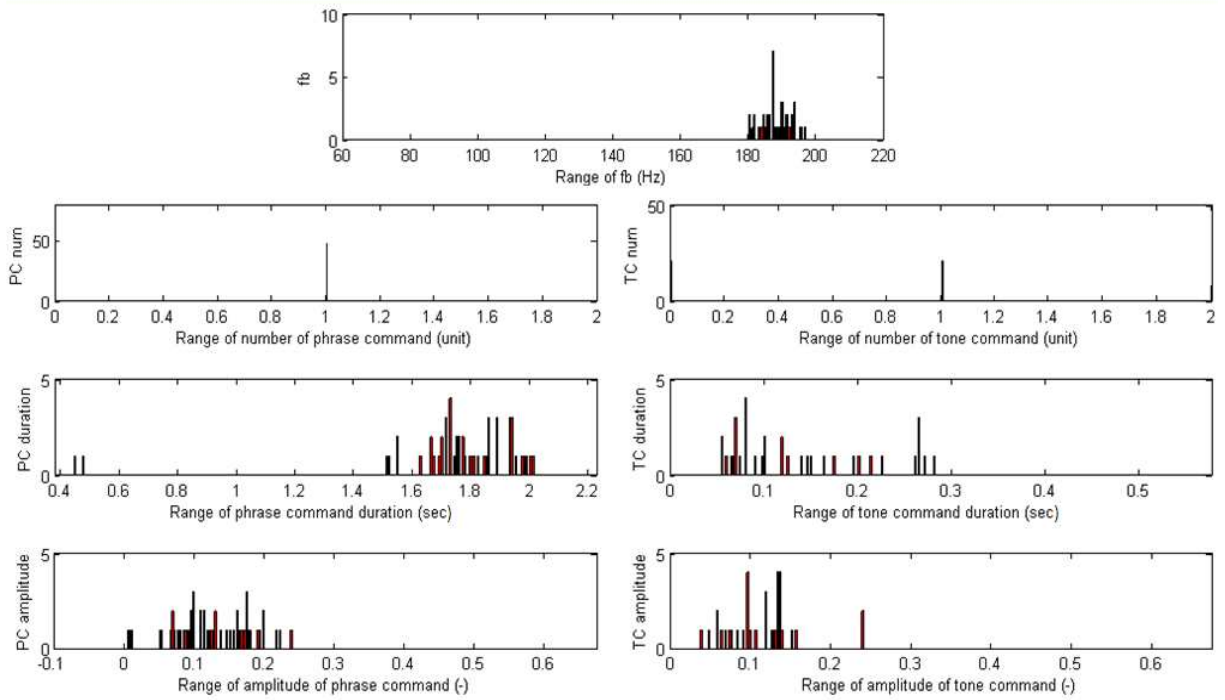Fig. 8: Female-speech frequency distributions of 7 parameters for tone 0

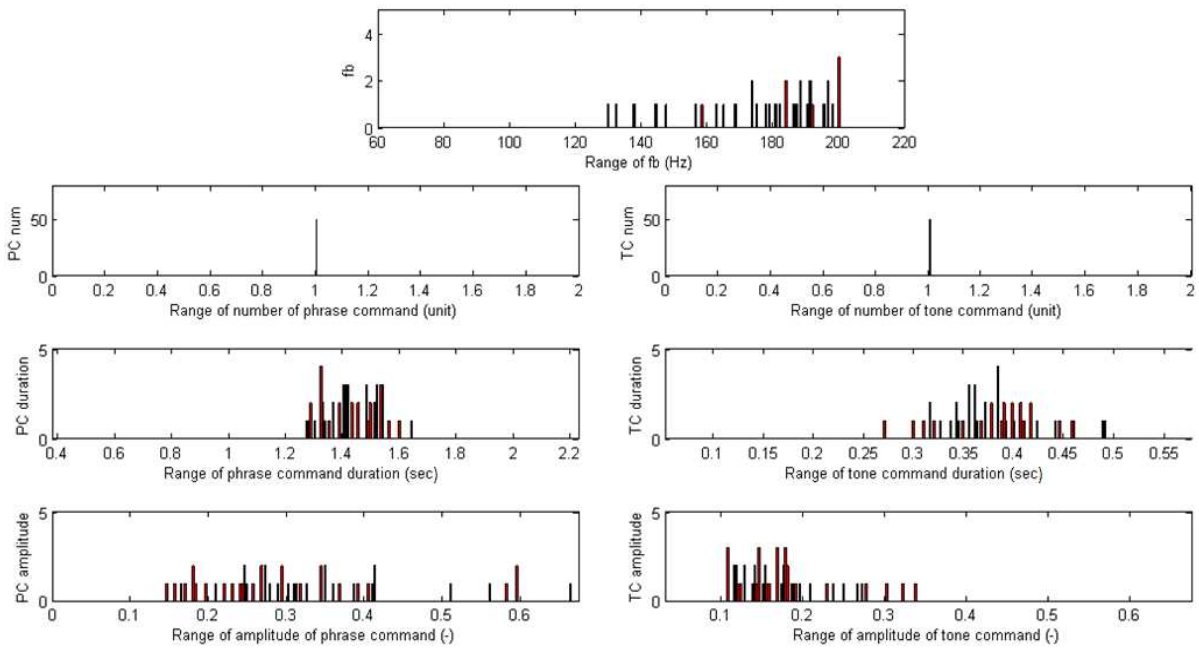Fig. 9: Female-speech frequency distributions of 7 parameters for tone 1

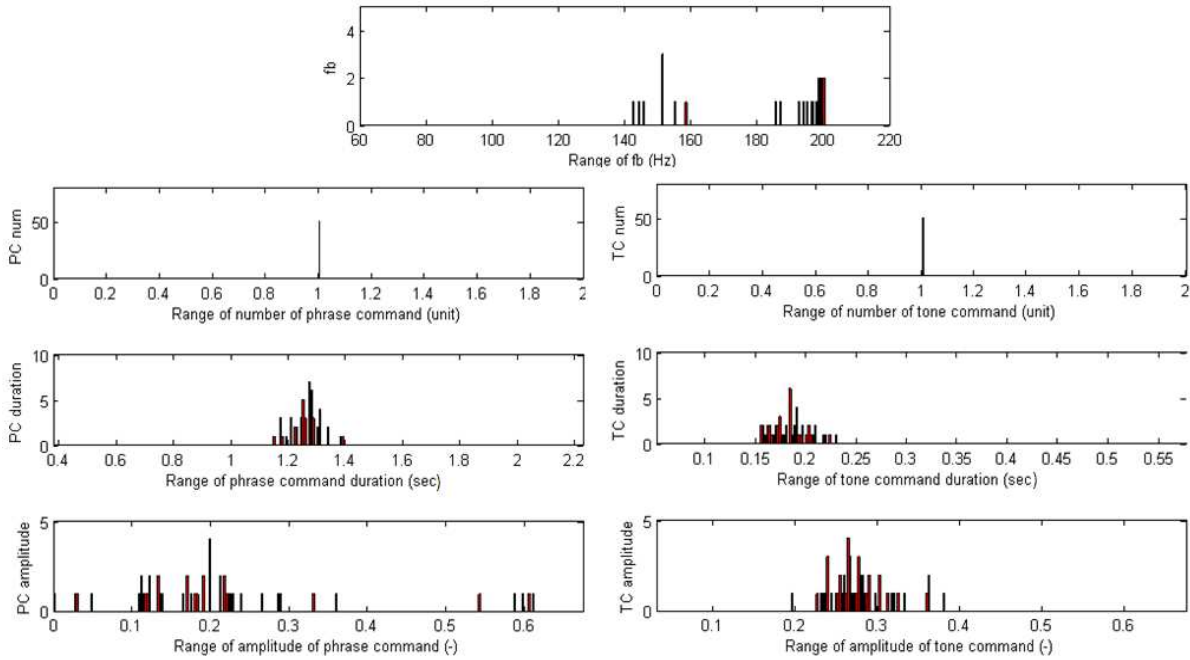Fig. 10: Female-speech frequency distributions of 7 parameters for tone 2

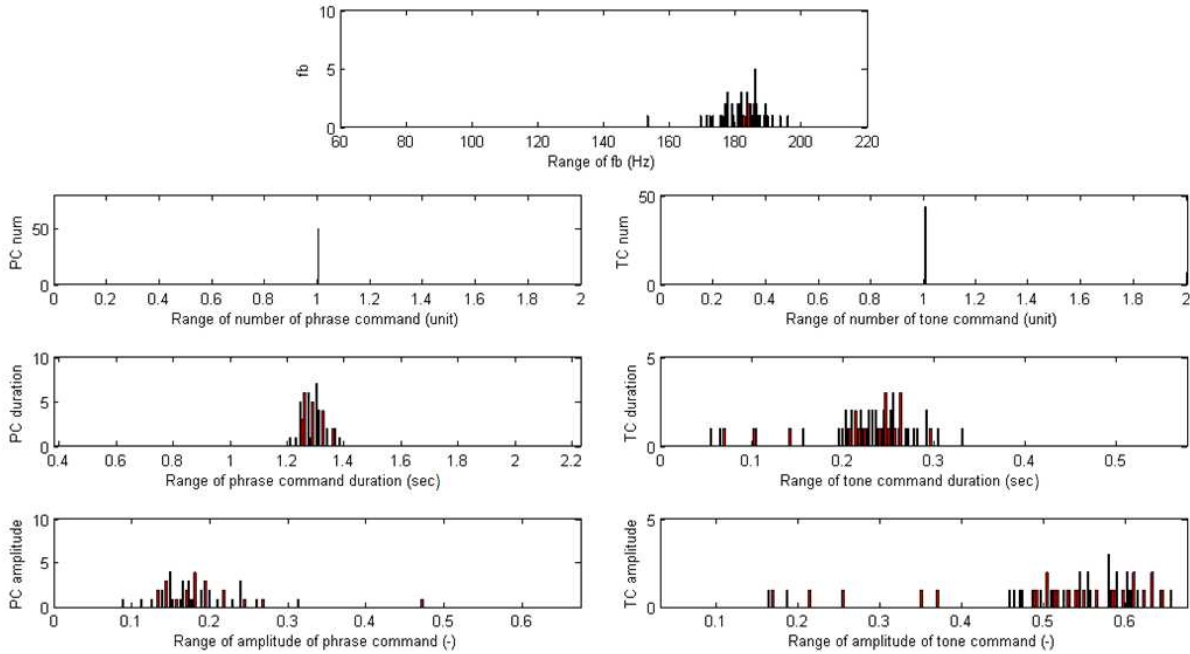Fig. 11: Female-speech frequency distributions of 7 parameters for tone 3

Fig. 12: Female-speech frequency distributions of 7 parameters for tone 4

Moreover, we can empirically noticed the parameters of number of phrase command and number of tone command in Fig. 3-12, most of them has only one unit, except for that of male-speech tone1in Fig. 4. Moreover, we can empirically noticed the parameter of phrase command duration in Fig. 3-12, mean value of tone 3 is the lowest for male speech, while the mean values of tone 3 and tone 4 are the about lowest for female speech. Considering the parameter of tone command duration in Fig. 3-12, we can see that

deviation of tone 0 is the highest for male speech, while the mean value of tone1 is the highest for female speech. Considering the parameter of amplitude of phrase command in Fig. 3-12, we can notice that the mean values for all five tones are quite similar for both male and female speech. Finally, considering the parameter of amplitude of tone command in Fig. 3-12, we can observe that the mean value of tone4 is the highest for both male and female speech.

CONCLUSION

Tone modeling is analyzed in this research. The speech material covers both man and woman speech and contains all fives tone in Thai language equally. The selected core modeling technique is of Fujisaki's. Most of seven derived ouput parameters are studied. The derived output parameter distributions discriminate one tone from the others.

ACKNOWLEDGEMENT

REFERENCES

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., 2011. Modeling of fundamental frequency contour of thai expressive speech using Fujisaki's model and structural model. J. Comput. Sci., 7: 1310-1317. DOI: 10.3844/jcssp.2011.1310.1317

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 a generative model of F0 contours for multilingual speech synthesis. Proceeding of the International Conference on Spoken Language Processing, (SLP' 98), IEEE Xpole 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. Proceeding of the International Conference on Spoken Language Processing, (SLP' 90), CiteULike, pp: 841-844.

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. 26-30, IEEE Xpole 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 Eurospeech, 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. Proceeding of the International Conference on Spoken Language Processing, Sept. 16-20, Denver, Colorado, USA., pp: 165-168.

Seresangtakul, P. and T. Takara, 2002. Analysis of pitch contour of thai tone using fujisaki's model. Proceeding of the International Conference on Acoustics, Speech and Signal Processing, May 13-17, IEEE Xpole 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. Proceeding of the International Conference on Acoustics, Speech and Signal Processing, Apr. 6-10, IEEE Xpole Press, 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. Proceeding of theTC-STAR Workshop on Speech-to-Speech Translation, Jun. 19-21, Barcelona, Spain, pp: 171-174.

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