

## Effects of Noises on Fujisaki's Model of Fundamental Frequency Contours for Thai Dialects

<sup>1,2</sup>Suphattharachai Chomphan and <sup>3</sup>Chutarat Chompunth

<sup>1</sup>Department of Electrical Engineering, Faculty of Engineering at Si Racha, Kasetsart University, 199 M.6, Tungsookhla, Si Racha, Chonburi, 20230, Thailand

<sup>2</sup>Center for Advanced Studies in Industrial Technology, Kasetsart University, 50 Ngam Wong Wan Rd, Ladyaow, Chatuchak, Bangkok, 10900, Thailand

<sup>3</sup>School of Social and Environmental Development, National Institute of Development Administration, 118 M.3, Serithai Road, Klong-Chan, Bangkok, 10240, Thailand

---

**Abstract: Problem statement:** Modeling of fundamental frequency (F0) contour plays an important role on the natural speech processing, since F0 is an important speech feature defining the human speech prosody. In Thai, there are four main dialects spoken by Thai people residing in four core region including central, north, northeast and south regions. Environmental noises are also plays an important role in corrupting the speech quality. The study of effects of noises on modeling of F0 contour for Thai dialects will evaluate robustness of the modeling techniques. **Approach:** The Fujisaki's model has been selected in this study because of its achievement in modeling of various Thai speech units. Four types of environmental noises are simulated with different levels of power. The differences among the model parameters of four Thai dialects have been summarized. This study proposes an analysis of model parameters for Thai speech prosody with four regional dialects and two genders and four types of noises. 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. **Results:** In the experiments, each regional dialect includes 10 samples of 10 sentences with male and female speech. Four types of noises include train, factory, car and air conditioner. Moreover, five levels of each type of noise are varied from 0-20 dB. The results show that most of the proposed parameters can distinguish four kinds of regional dialects explicitly. **Conclusion:** By using the Fujisaki's model, the results confirm that the proposed parameters can distinguish the regional dialects efficiently. However, the simulated noises deteriorate the F0 contours and also distort the model parameters.

**Key words:** Fundamental frequency modeling, environmental noise, fundamental frequency, command durations, phrase commands

---

### INTRODUCTION

The modeling of F0 contour with noisy environment causes the degradation of naturalness of the speech. To develop the natural speech processing system, it is necessary to know how the noise deteriorates the model parameters. Fujisaki's modeling of fundamental frequency for Thai expressive speech conducted in 2010 is proved to be effective for a limited-domain speech corpus (Chomphan, 2010a). It has been found that the

derived parameters can distinguish one style of speech from each other.

Fujisaki's Modeling of F0 contours for Thai Dialects has been conducted by (Chomphan 2010a; 2010b). However, the effects of noises have not been studied.

By using the same way of Thai dialects without considering of various types of noises (Chomphan, 2010b), the study proposes an analysis of F0 modeling of four Thai dialects including standard Thai, Lanna or North

---

**Corresponding Author:** Suphattharachai Chomphan, Department of Electrical Engineering, Faculty of Engineering at Si Racha, Kasetsart University, 199 M.6, Tungsookhla, Si Racha, Chonburi, 20230, Thailand

dialect, Lao-style or North East dialect and South dialect with four different types of noises. 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.

### MATERIALS AND METHODS

**Fujisaki's model:** The F0 contour of an utterance of speech is treated as a linear superposition of a global phrase and local accent components on a logarithmic scale, as depicted in Fig. 1 (Fujisaki and Sudo, 1971). By using this generative model, the parameters are extracted from our speech database, utterance by utterance. Subsequently, the derived parameters are computed and analyzed (Chomphan and Kobayashi, 2008; 2009; Seresantakul and Takara, 2003).

**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.

**Environmental noises:** Four types of noises include train, factory, car and air conditioner.

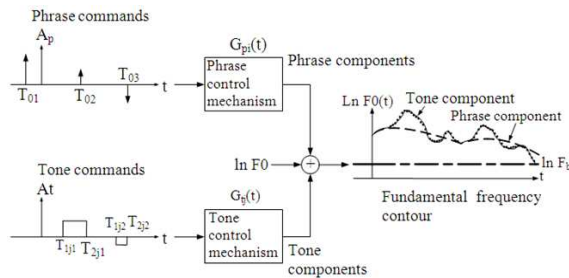


Fig. 1: An extension of Fujisaki's model for the generation of F0 contour

They are mixed directly with the pre-recorded clean speech in the speech database. Before mixing noises with the clean speech, the noise volume or power are adjusted in several exact levels. As for the level variation of noises, the levels of each type of noise are varied from 0, 5, 10, 15, 20 dB, respectively.

### RESULTS

In our speech database, we use ten 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 a hundred utterances of samples. Therefore we have four hundred 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 are plot in a graph to show the differences and similarities among those dialects. The first five graphs are of female Center dialect (Fig. 2-6), the second five graphs are of female North dialect (Fig. 7-11), the next five graphs are of female Northeast dialect (Fig. 12-16), while the next five ones are of female South dialect (Fig. 17-21). The following twenty graphs (Fig. 22-41) are of male speech with the same order as that of female speech. 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. Please be noted that the first sub-graph in the following figures is not the main list of 7 model parameters. However it reflects the distribution of length of utterance.

In each dialect, there are five representative graphs which are the parameter distribution of clean speech, air-conditioner noise corrupted speech, car noise corrupted speech, factory noise corrupted speech and train noise corrupted speech, respectively.

### DISCUSSION

From the frequency distribution graphs of female and male speech in Fig. 2-41, most results show that the distributions of four dialects are significantly different.

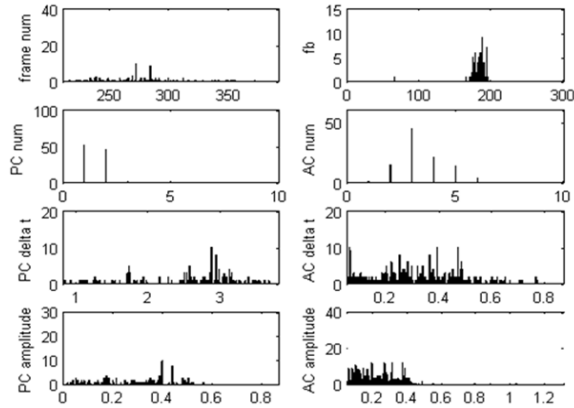


Fig. 2: Comparison of 7-parameter distributions of female center-dialect clean speech

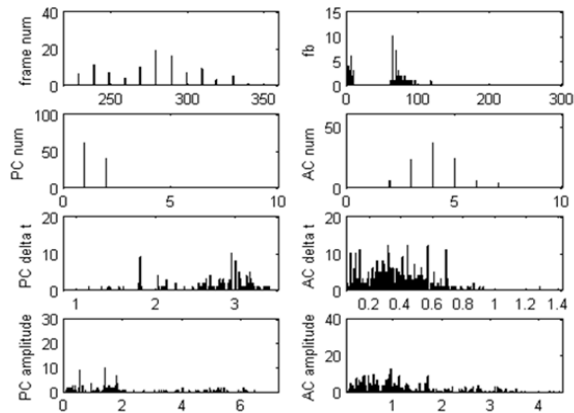


Fig. 3: Comparison of 7-parameter distributions of female center-dialect air-conditioner noise corrupted speech

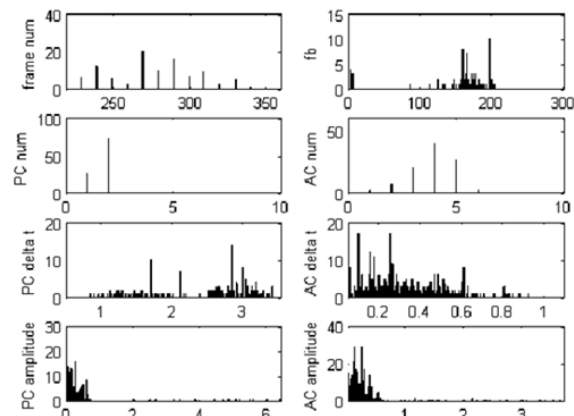


Fig. 4: Comparison of 7-parameter distributions of female center-dialect car noise corrupted speech

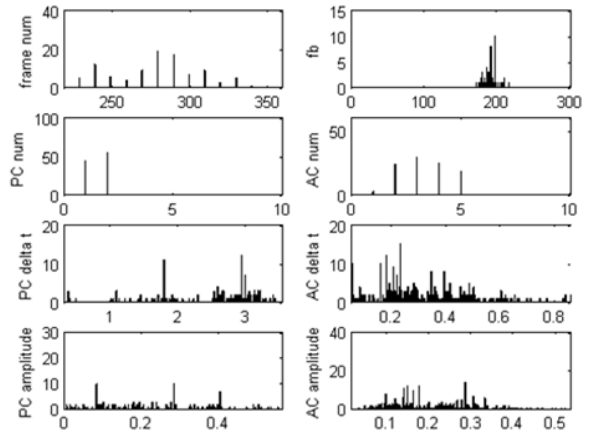


Fig. 5: Comparison of 7-parameter distributions of female center-dialect factory noise corrupted speech

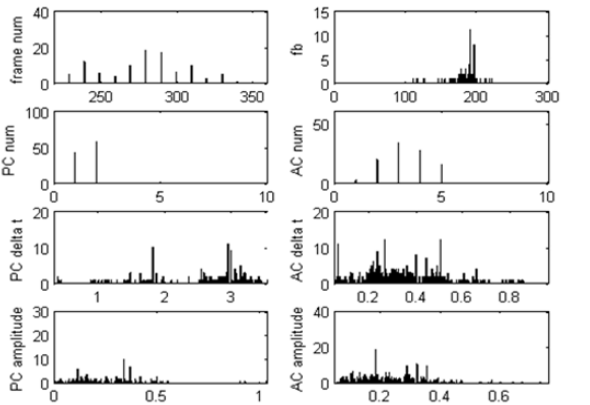


Fig. 6: Comparison of 7-parameter distributions of female center-dialect train noise corrupted speech

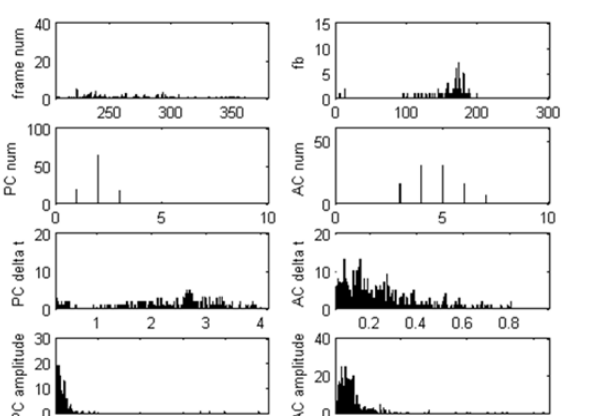


Fig. 7: Comparison of 7-parameter distributions of female north-dialect clean speech

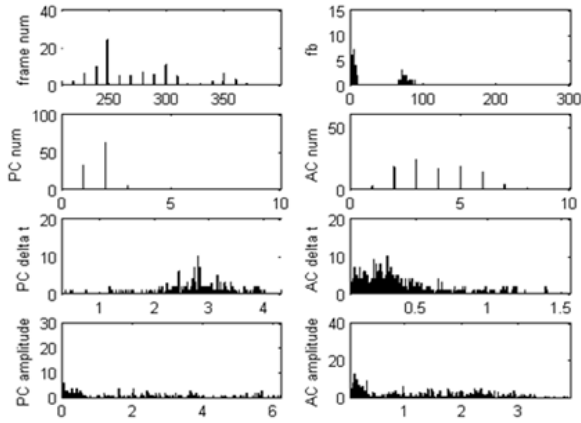


Fig. 8: Comparison of 7-parameter distributions of female north-dialect air-conditioner noise corrupted speech

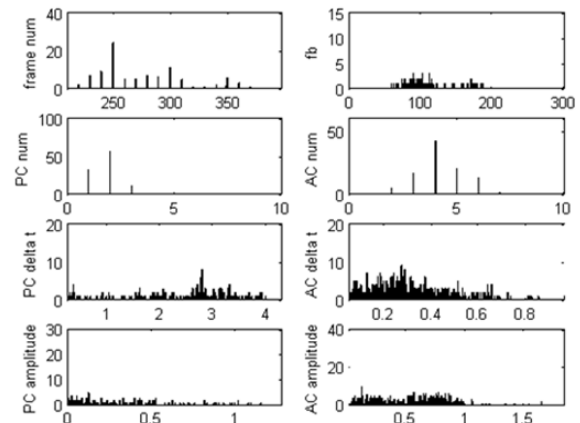


Fig. 11: Comparison of 7-parameter distributions of female north-dialect train noise corrupted speech

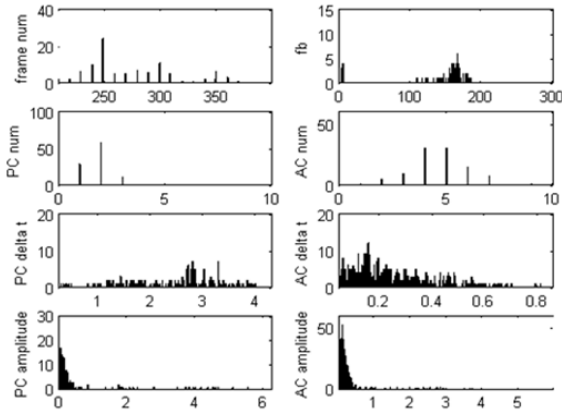


Fig. 9: Comparison of 7-parameter distributions of female north-dialect car noise corrupted speech

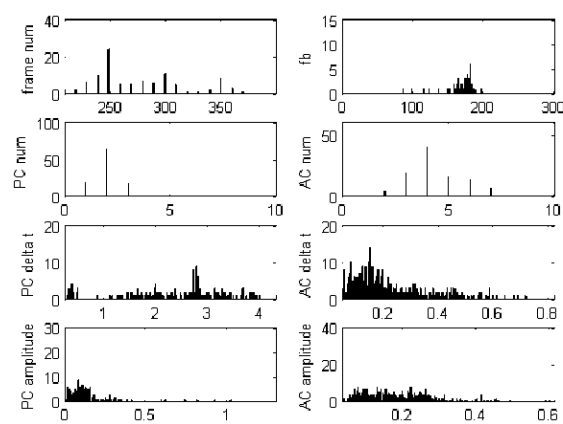


Fig. 12: Comparison of 7-parameter distributions of female northeast-dialect clean speech

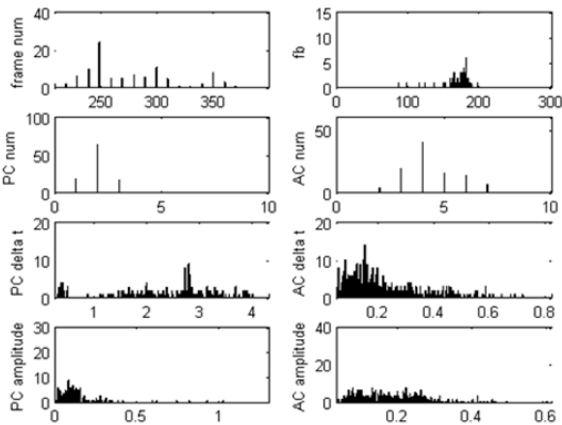


Fig. 10: Comparison of 7-parameter distributions of female north-dialect factory noise corrupted speech

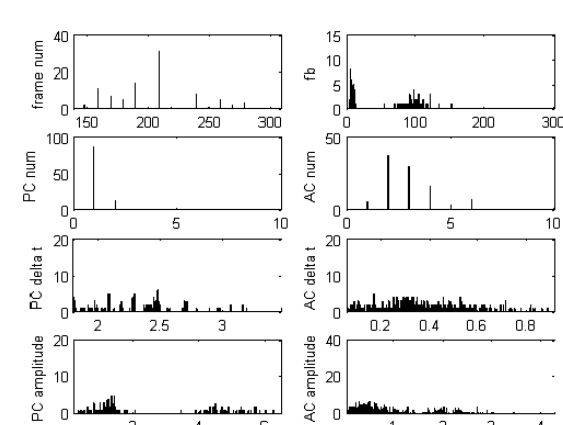


Fig. 13: Comparison of 7-parameter distributions of female northeast-dialect air-conditioner noise corrupted speech

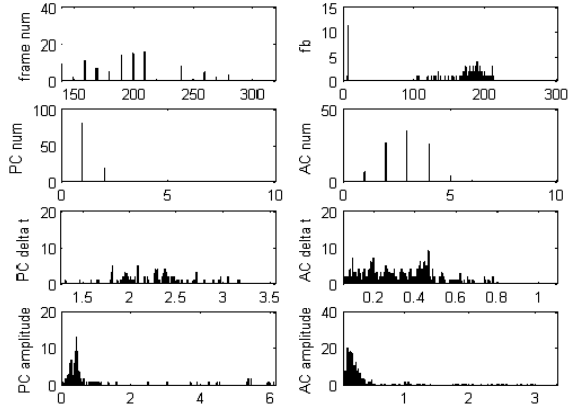


Fig. 14: Comparison of 7-parameter distributions of female northeast-dialect car noise corrupted speech

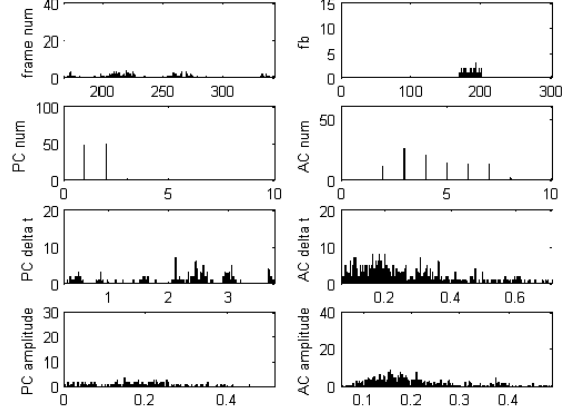


Fig. 17: Comparison of 7-parameter distributions of female south-dialect clean speech

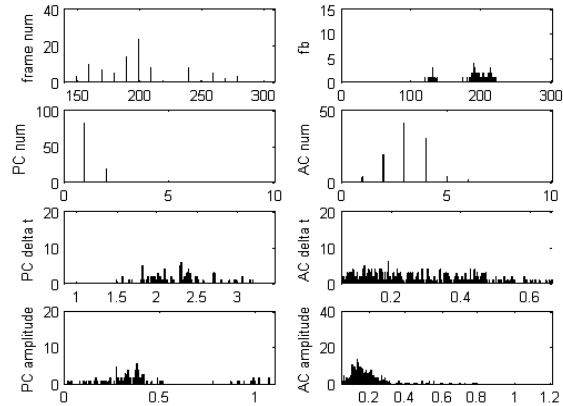


Fig. 15: Comparison of 7-parameter distributions of female northeast-dialect factory noise corrupted speech

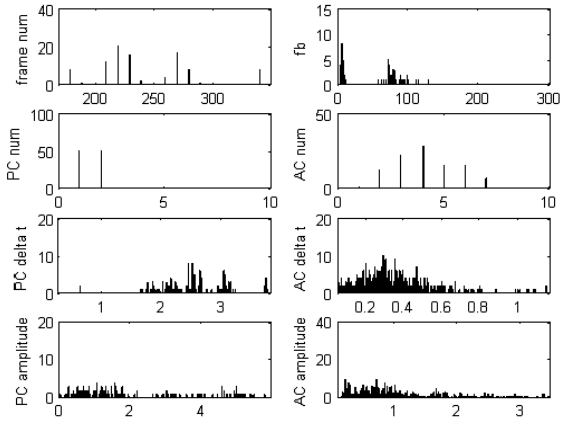


Fig. 18: Comparison of 7-parameter distributions of female south-dialect air-conditioner noise corrupted speech

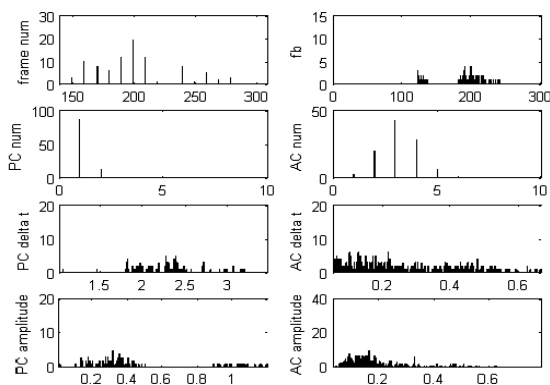


Fig. 16: Comparison of 7-parameter distributions of female northeast-dialect train noise corrupted speech

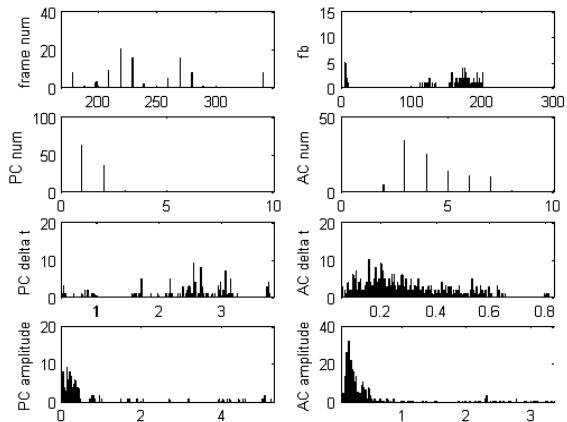


Fig. 19: Comparison of 7-parameter distributions of female south-dialect car noise corrupted speech

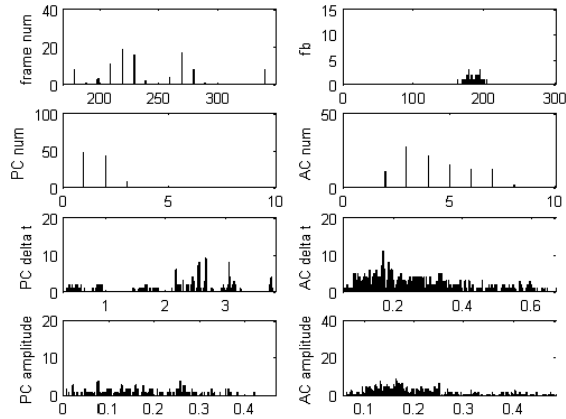


Fig. 20: Comparison of 7-parameter distributions of female south-dialect factory noise corrupted speech

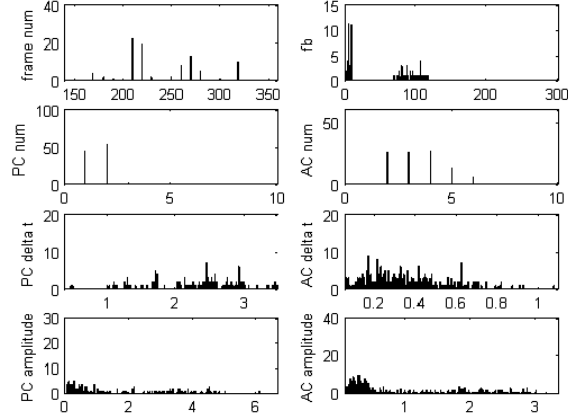


Fig. 23: Comparison of 7-parameter distributions of male center-dialect air-conditioner noise corrupted speech

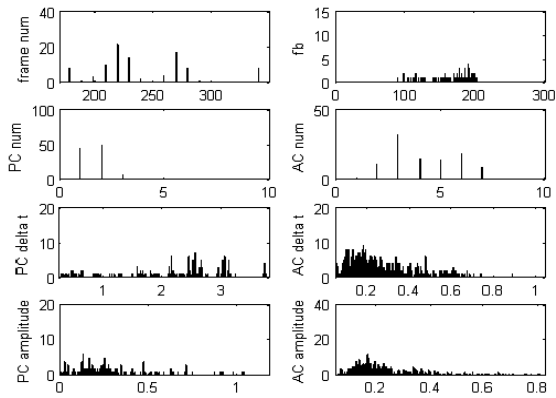


Fig. 21: Comparison of 7-parameter distributions of female south-dialect train noise corrupted speech

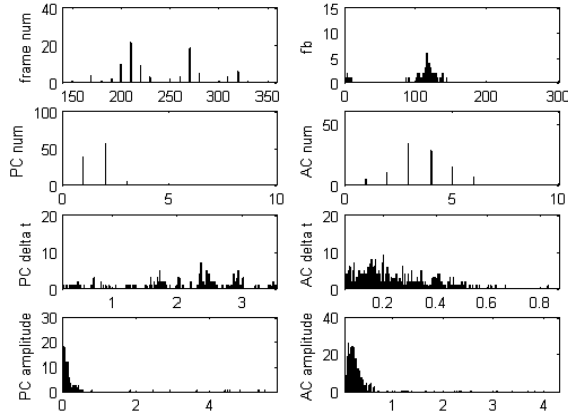


Fig. 24: Comparison of 7-parameter distributions of male center-dialect car noise corrupted speech

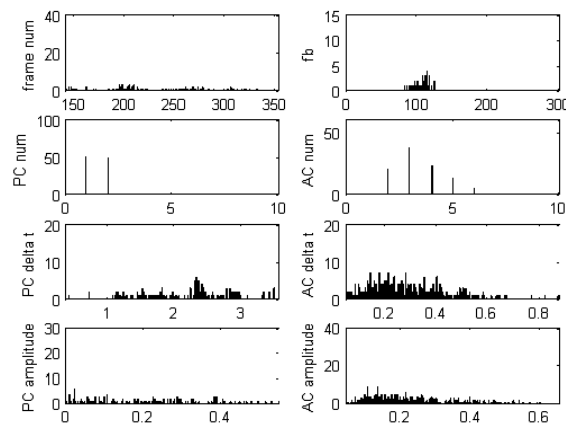


Fig. 22: Comparison of 7-parameter distributions of male center-dialect clean speech

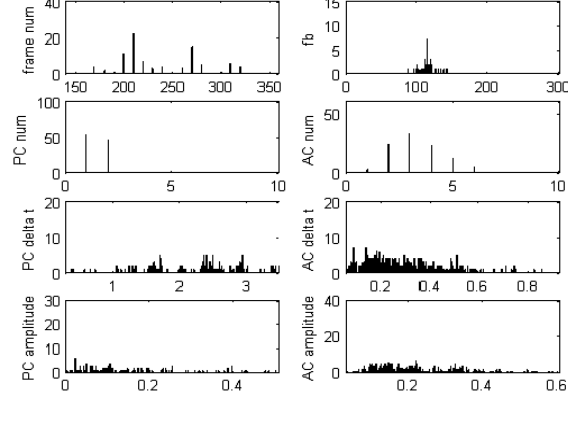


Fig. 25: Comparison of 7-parameter distributions of male center-dialect factory noise corrupted speech

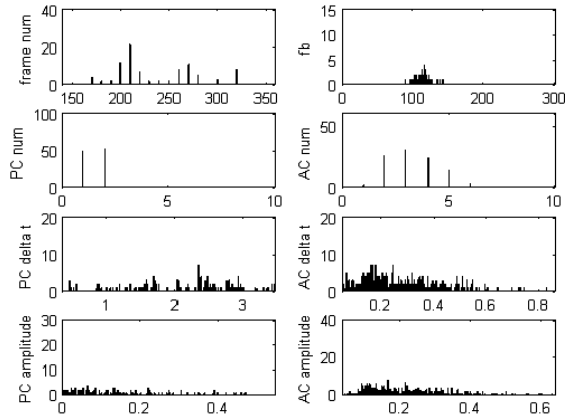


Fig. 26: Comparison of 7-parameter distributions of male center-dialect train noise corrupted speech

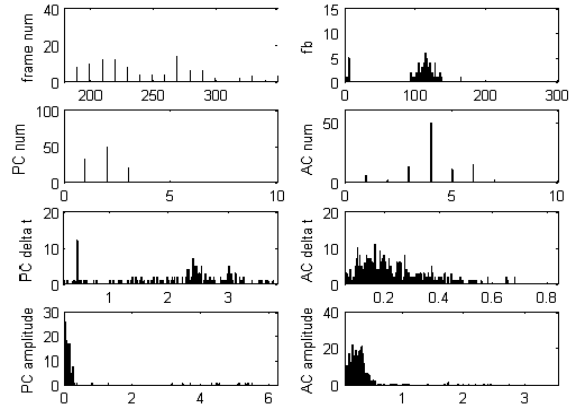


Fig. 29: Comparison of 7-parameter distributions of male north-dialect car noise corrupted speech

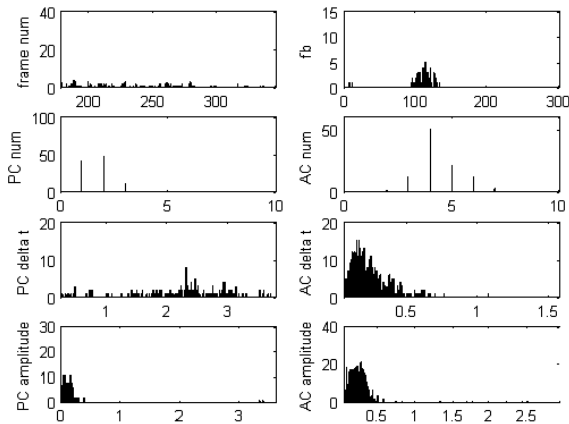


Fig. 27: Comparison of 7-parameter distributions of male north-dialect clean speech

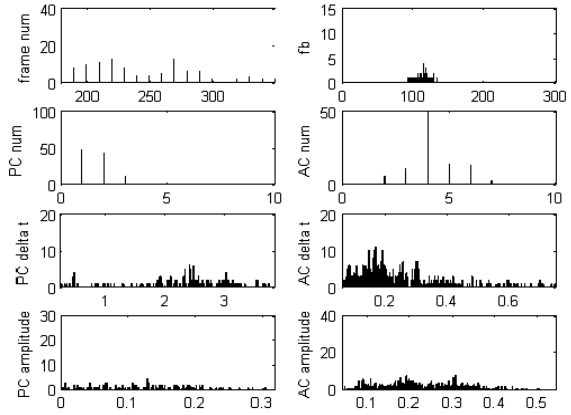


Fig. 30: Comparison of 7-parameter distributions of male north-dialect factory noise corrupted speech

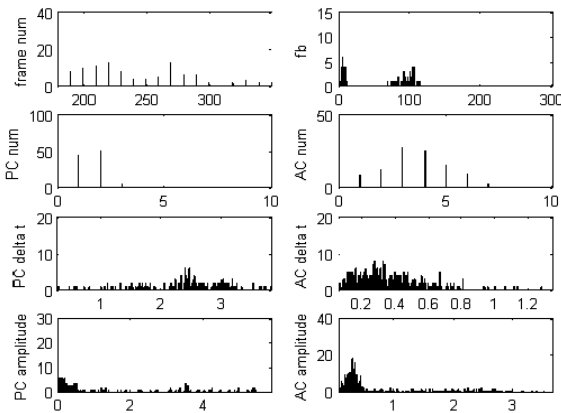


Fig. 28: Comparison of 7-parameter distributions of male north-dialect air-conditioner noise corrupted speech

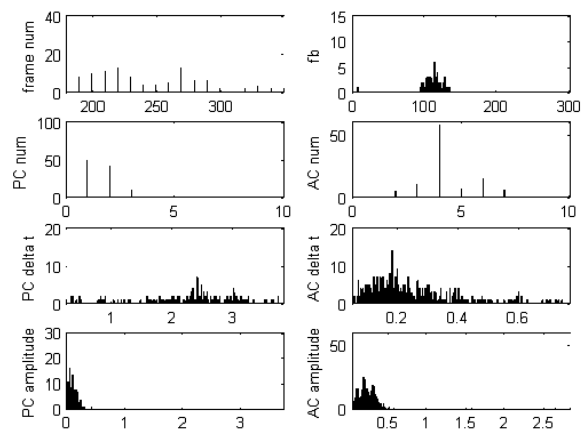


Fig. 31: Comparison of 7-parameter distributions of male north-dialect train noise corrupted speech

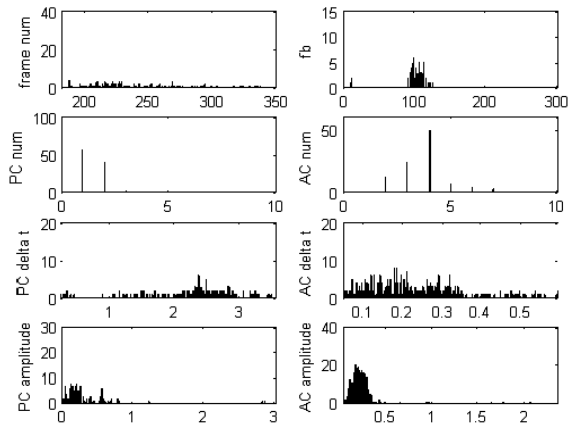


Fig. 32: Comparison of 7-parameter distributions of male northeast-dialect clean speech

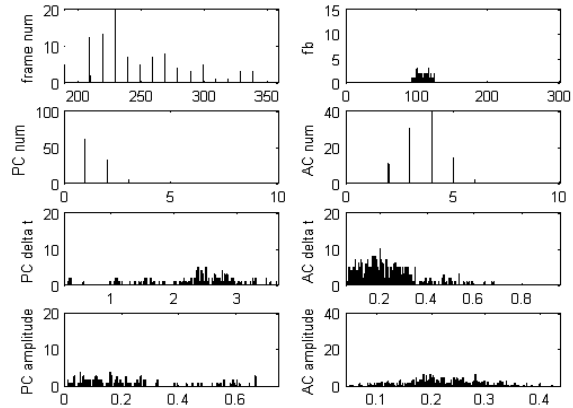


Fig. 35: Comparison of 7-parameter distributions of male northeast-dialect factory noise corrupted speech

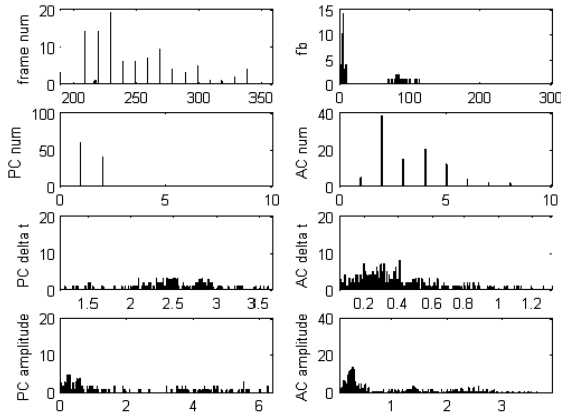


Fig. 33: Comparison of 7-parameter distributions of male northeast-dialect air-conditioner noise corrupted speech

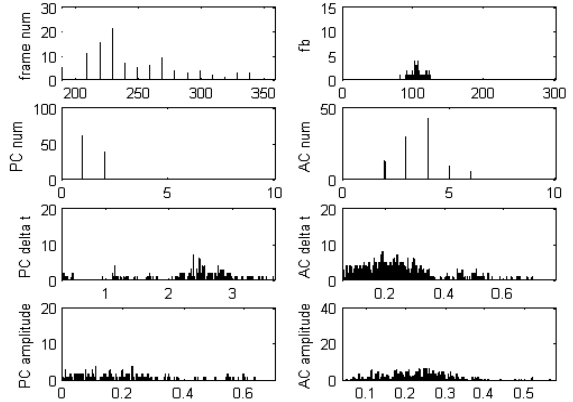


Fig. 36: Comparison of 7-parameter distributions of male northeast-dialect train noise corrupted speech

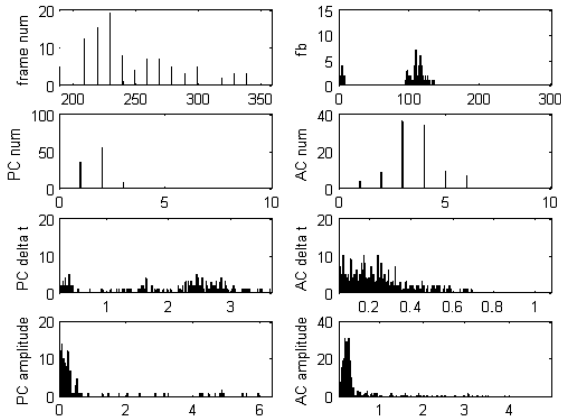


Fig. 34: Comparison of 7-parameter distributions of male northeast-dialect car noise corrupted speech

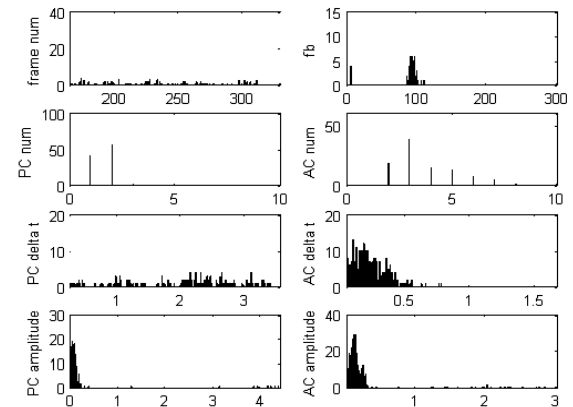


Fig. 37: Comparison of 7-parameter distributions of male south-dialect clean speech



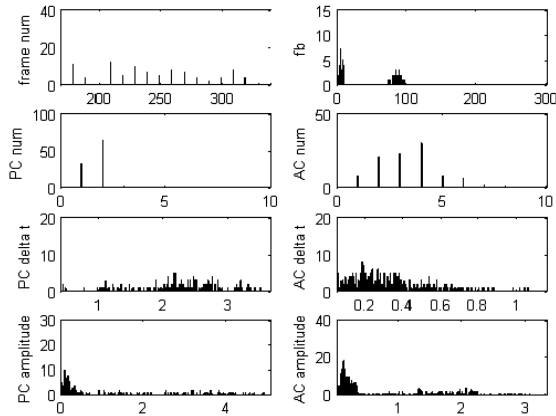


Fig. 38: Comparison of 7-parameter distributions of male south-dialect air-conditioner noise corrupted speech

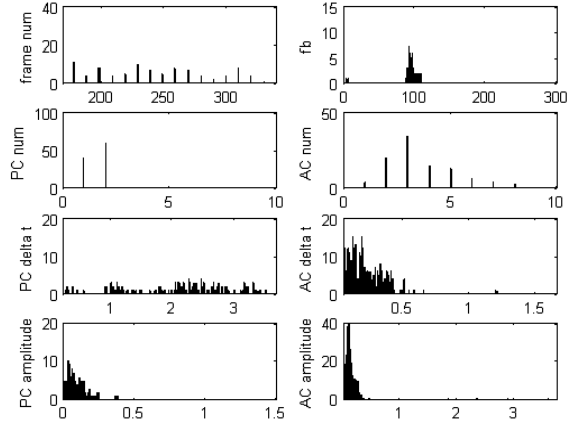


Fig. 41: Comparison of 7-parameter distributions of male south-dialect train noise corrupted speech

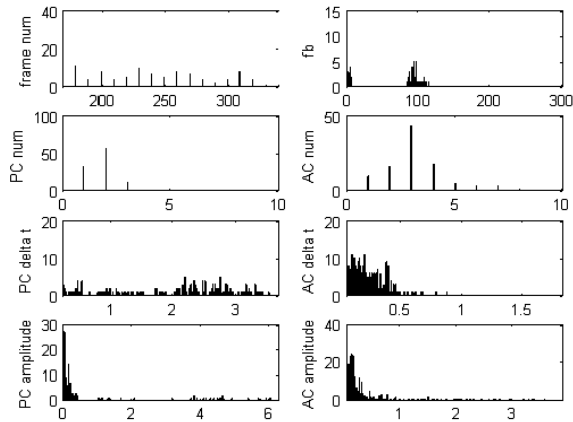


Fig. 39: Comparison of 7-parameter distributions of male south-dialect car noise corrupted speech

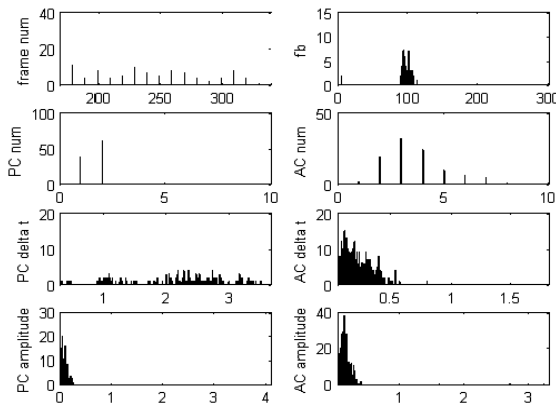


Fig. 40: Comparison of 7-parameter distributions of male south-dialect factory noise corrupted speech

As for most of female speech in Fig. 2-21, it can be empirically seen that the air-conditioner and car noises cause their distributions of amplitude of phrase command and amplitude of tone command become more fragmented significantly. This observation can also be seen for male speech in Fig. 22-41. Moreover, it can be noticed that the air-conditioner noise cause the center of the corresponding distribution of baseline frequency reduces to half approximately. However this observation cannot be seen for male speech. The level of noise selected to be presented is of 10 dB. It has been noted that the other levels of noises do have the corresponding results.

### CONCLUSION

This study presents a study of effects of noises on modeling of F0 contour for Thai dialects. The Fujisaki's model has been selected in this study. Four types of environmental noises are simulated with different levels of power. The differences among the model parameters of four Thai dialects have been summarized. The experimental results show that most of the proposed parameters can distinguish four kinds of regional dialects explicitly. Some types of noises cause some differences in the distributions of the model parameters. All in all, the simulated noises deteriorate the F0 contours and also distort the model parameters.

### ACKNOWLEDGEMENT

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

**REFERENCES**

- Chomphan, S. and T. Kobayashi, 2007a. Design of tree-based context clustering for an HMM-based Thai speech synthesis system. Proceeding 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. 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., 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
- Fujisaki, H. and H. Sudo, 1971. A model for the generation of fundamental frequency contours of Japanese word accent. *J. Acoust. Soc. Jap.*, 57: 445-452.
- Mixdorff, H. and H. Fujisaki, 1997. Automated quantitative analysis of F0 contours of utterances from a German ToBI-labeled speech database. Proceeding of the Eurospeech, Sept. 22-25, Rhodes, Greece, pp: 187-190.
- 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 Xplore Press, Hong Kong, pp: 452-455.