American Journal of Applied Sciences 9 (8): 1251-1258, 2012
ISSN 1546-9239

Fujisaki's Model of Thai's Fundamental Frequency Contours with Environmental Noises

[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: An important human speech feature is the fundamental frequency (F0) contour which represents the speech prosody. It indicates the naturalness and intelligibility of the speech. Modeling of fundamental frequency contour was an essential procedure in the natural speech processing. In speech communication, environmental noise plays an essential role in damaging the digital communication quality. The study of effects of noises on modeling of F0 contour for standard Thai is conducted. **Approach:** The selected modeling technique in this study was adapted from Fujisaki's model, because of its achievement in modeling of various Thai speech units. Four types of environmental noises were recorded for different levels of power. This study was proposed an analysis of some parameters of modeling of Thai speech prosody for two genders and four types of noises. The derived Fujisaki's model was covered seven parameters including 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, the standard Thai of 2 samples of 5 sentences with 5 males and 5 females was used. Four types of noises include train, factory, car and air conditioner. Five levels of each type of noise were varied from 0-20 dB. The results were showing that the different noises give the distinguished effects for most of the proposed model parameters. **Conclusion:** The results confirm that the effects of four types of noises are significantly different. It can be seen that the environmental noises deteriorate the model parameters empirically.

Key words: Standard Thai, Fujisaki's model, analysis of fundamental frequency, environmental noise, noise effect, fundamental frequency contours

INTRODUCTION

In the former study, modeling of F0 contour with noisy environment causes the deterioration of naturalness of the speech. To develop the modern natural speech processing system, it is very important to know how the noise degrades the model parameters. The previous study of F0 modeling has been conducted in many levels of speech units, for examples, utterance level, word and syllable levels (Fujisaki and Sudo, 1971; Fujisaki *et al*., 1990; Fujisaki and Ohno, 1998; Saito and Sakamoto, 2002; Li *et al*., 2004; Tao *et al*., 2006; Ni and Hirose, 2006; Tran *et al*., 2006). Moreover, in Thai speech, this model has been effectively applied for applying to the utterances, words and tones (Seresangtakul and Takara, 2002; 2003; Hiroya and Sumio, 2002). The modeling of fundamental frequency for Thai expressive speech with a limited-domain speech database was succesfully conducted in 2010 (Chomphan, 2010a). It has been seen that the selected model parameters are able to distinguish all styles of expressive speech.

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

This study applies the same way of the former study by using an analysis of F0 contour modeling of standard Thai with four different types of noises. The Fujisaki's model is a basic tool for applying in the the advanced research of the natural speech recognition and synthesis. (Seresangtakul and Takara, 2002; 2003; Chomphan, 2010c; 2011a).

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

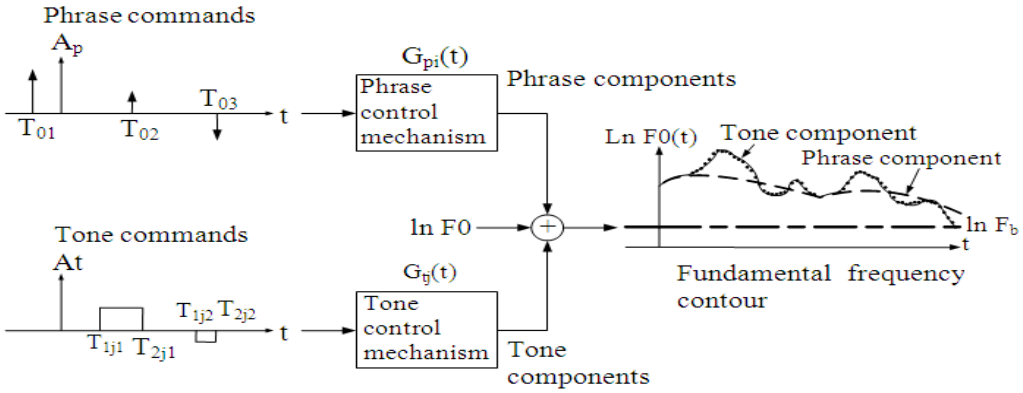

Fig. 1: An extension system of Fujisaki's model

MATERIALS AND METHODS

Fujisaki's model: Fig. 1 illustrates the F0 contour of an utterance of speech which is treated as a linear superposition of a local accent component and a global phrase component on a logarithmic scale. (Fujisaki and Sudo, 1971). By appling the Fujisaki's model, the related parameters are extracted from the speech corpus, utterance by utterance. Therafter the output parameters are calculated are systematically analyzed (Chomphan and Kobayashi, 2008; 2009; Seresangtakul and Takara, 2003).

Derived parameters: Seven derived parameters are calculated from the conventional parameters. It is noted that these derived paramenters mostly reflect the geometrical appearance of the F0 contour of the speech:

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

The derived parameters are mostly extracted for Thai speech. It has been noted that number of frame is also extracted in the experimental results. However it is not the main focused parameters explained earlier (Chomphan, 2011b).

Environmental noises: To evaluate the effects of noises in speech communication, four types of noises including train, factory, car and air conditioner are recorded. 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 the speech corpus, we use standard Thai of 2 samples of 5 sentences with 5 males and 5 females. The sentences have been recorded in standard Thai. Both male and female speech has been constructed in the speech database. The extraction tools are applied in this study (Mixdorff and Fujisaki, 1997; Chomphan and Kobayashi, 2007a; 2007b). For each parameter, the frequency distribution over its range is constructed, subsequently the distributions of standard Thai are plot in a graph. The differences and similarities among those different types of noises are illustrated in the graph. The first graph is of clean speech (Fig. 2), Fig. 3 and 4 are of speech corrupted by air-conditioner noises at 0 and 20 dB, respectively. Figure 5 and 6 are of speech corrupted by car noises at 0 and 20 dB, respectively. Figure 7 and 8 are of speech corrupted by factory noises at 0 and 20 dB, respectively. Figure 9 and 10 are of speech corrupted by train noises at 0 and 20 dB, respectively. These abbreviations are used in most figures; frame num, fb, AC num, PC num, AC delta t, PC delta t, AC amplitude and PC amplitude, mean number of frames, baseline frequency, number of tone commands, number of phrase commands, tone command duration, phrase command duration, amplitude of tone command and amplitude of phrase 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.

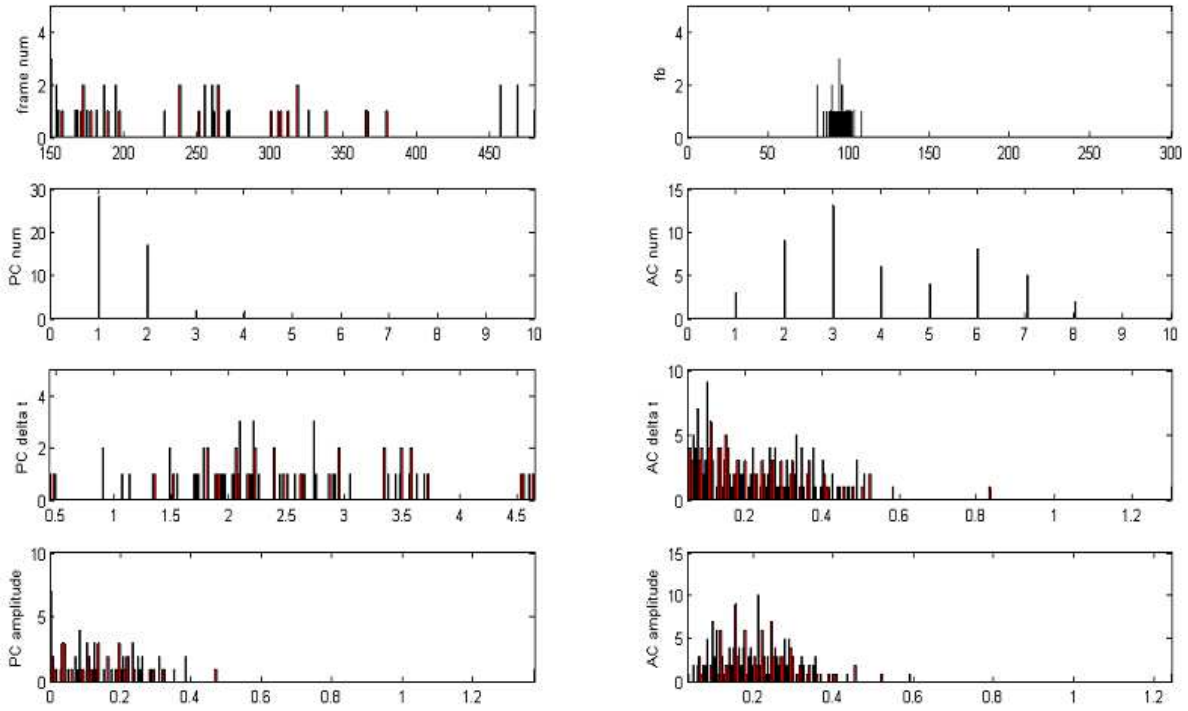

Fig. 2: Comparison of 7-parameter distributions of standard Thai clean speech

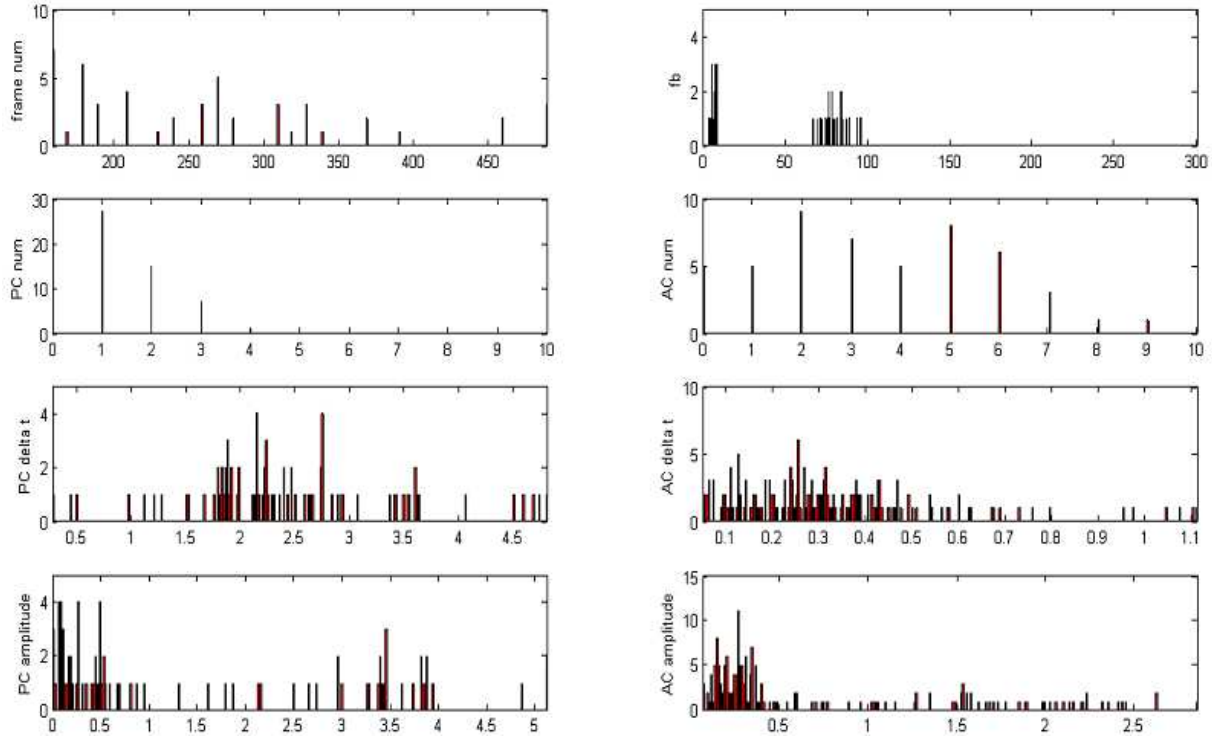

Fig. 3: Comparison of 7-parameter distributions of standard Thai air-conditioner noise corrupted speech at 0 dB

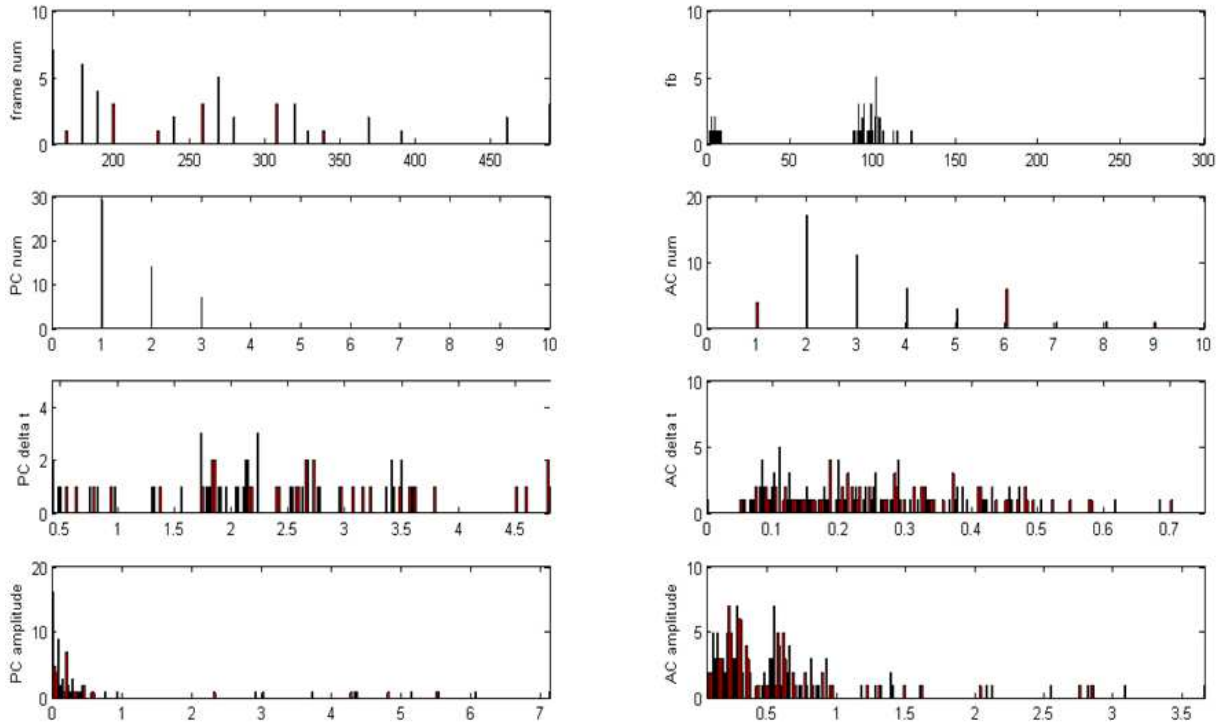

Fig. 4: Comparison of 7-parameter distributions of standard Thai air-conditioner noise corrupted speech at 20 dB

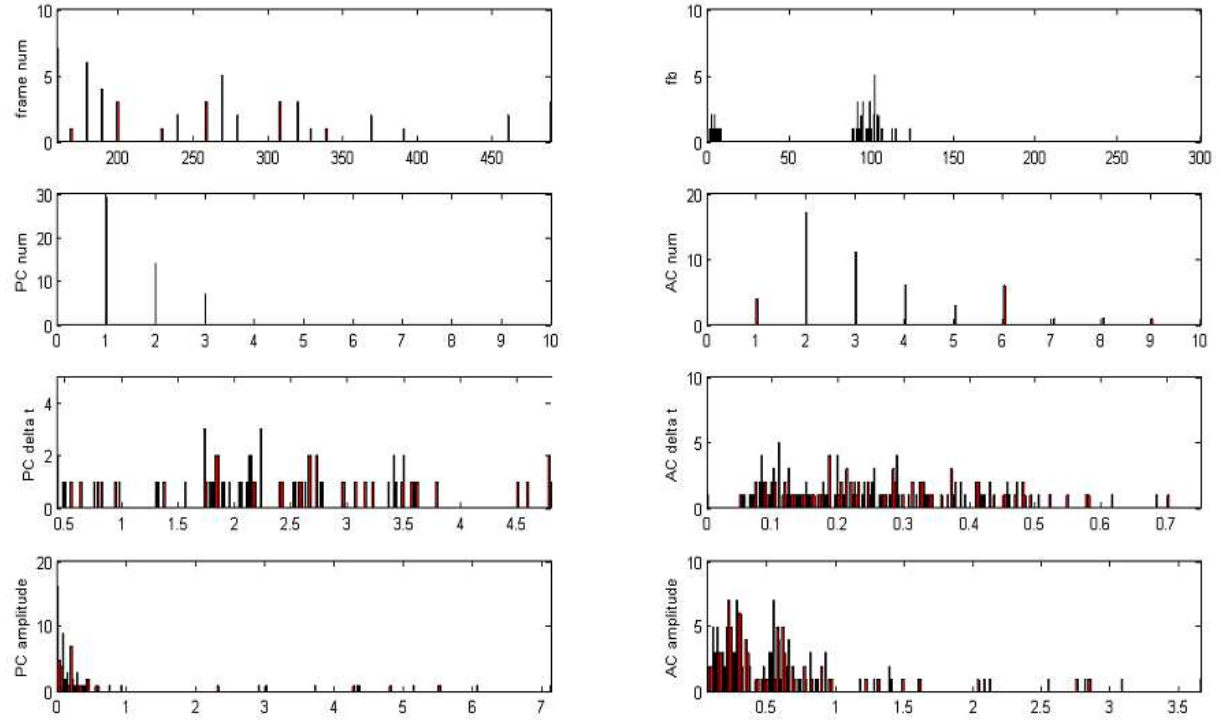

Fig. 5: Comparison of 7-parameter distributions of standard Thai car noise corrupted speech at 0 dB

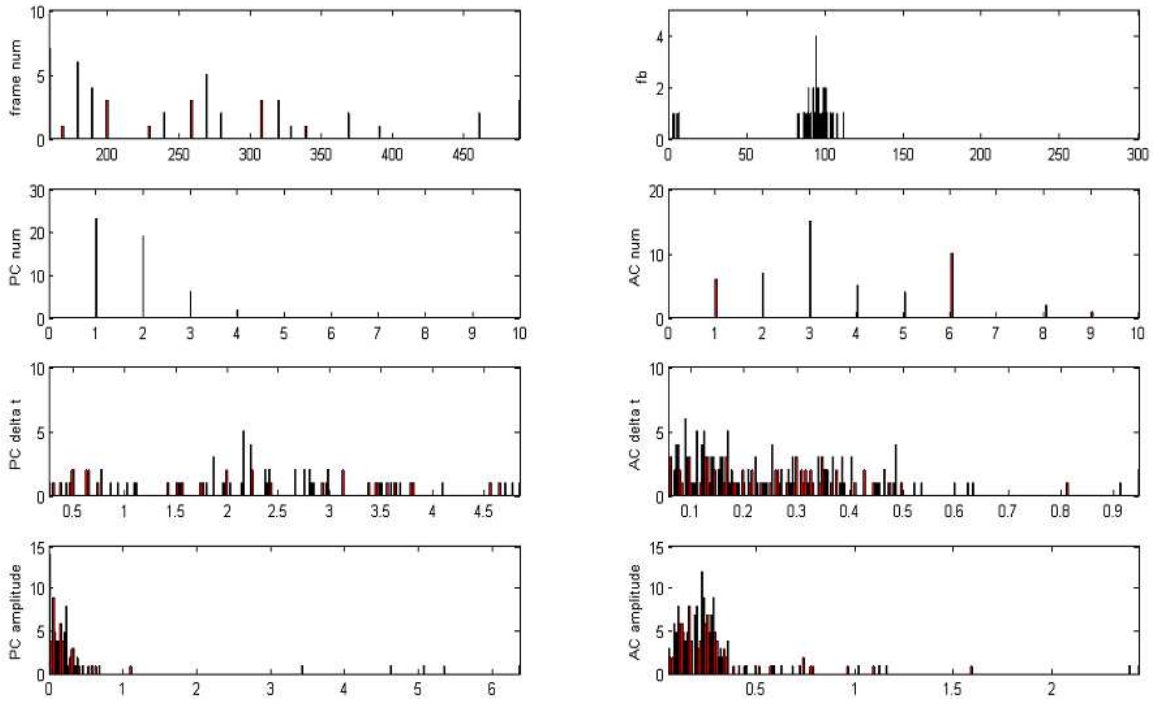

Fig. 6: Comparison of 7-parameter distributions of standard Thai car noise corrupted speech at 20 dB

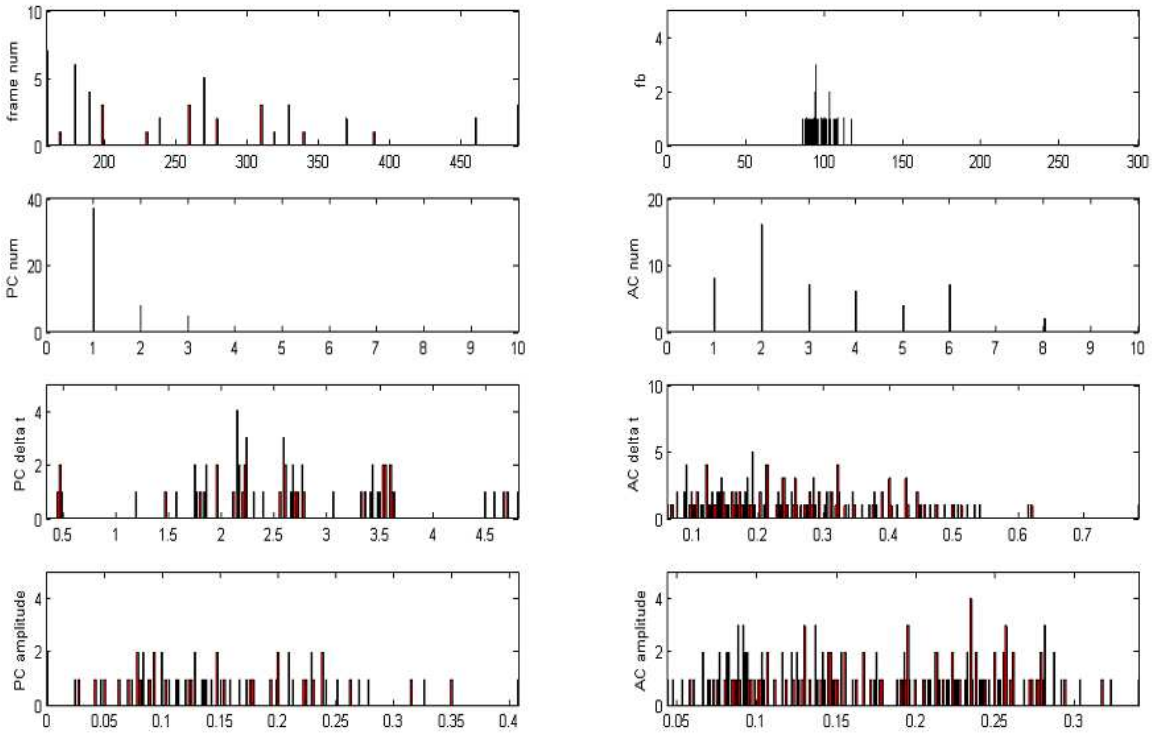

Fig. 7: Comparison of 7-parameter distributions of standard Thai factory noise corrupted speech at 0 dB

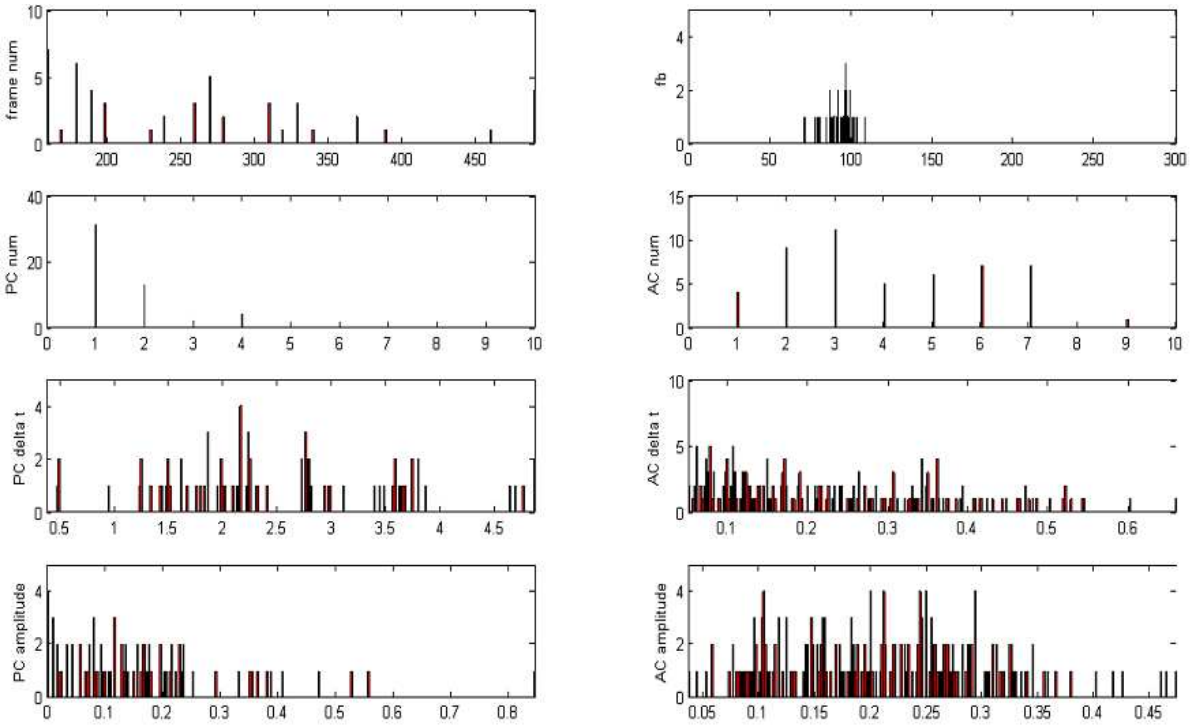

Fig. 8: Comparison of 7-parameter distributions of standard Thai factory noise corrupted speech at 20 dB

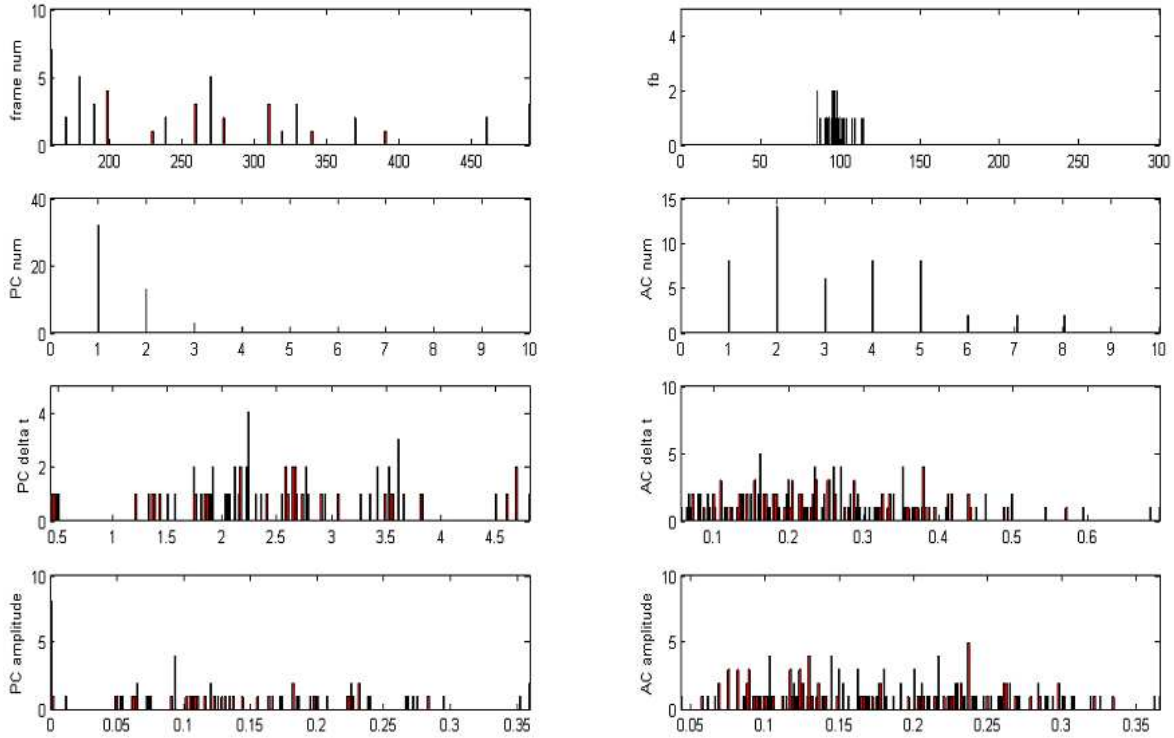

Fig. 9: Comparison of 7-parameter distributions of standard Thai train noise corrupted speech at 0 dB

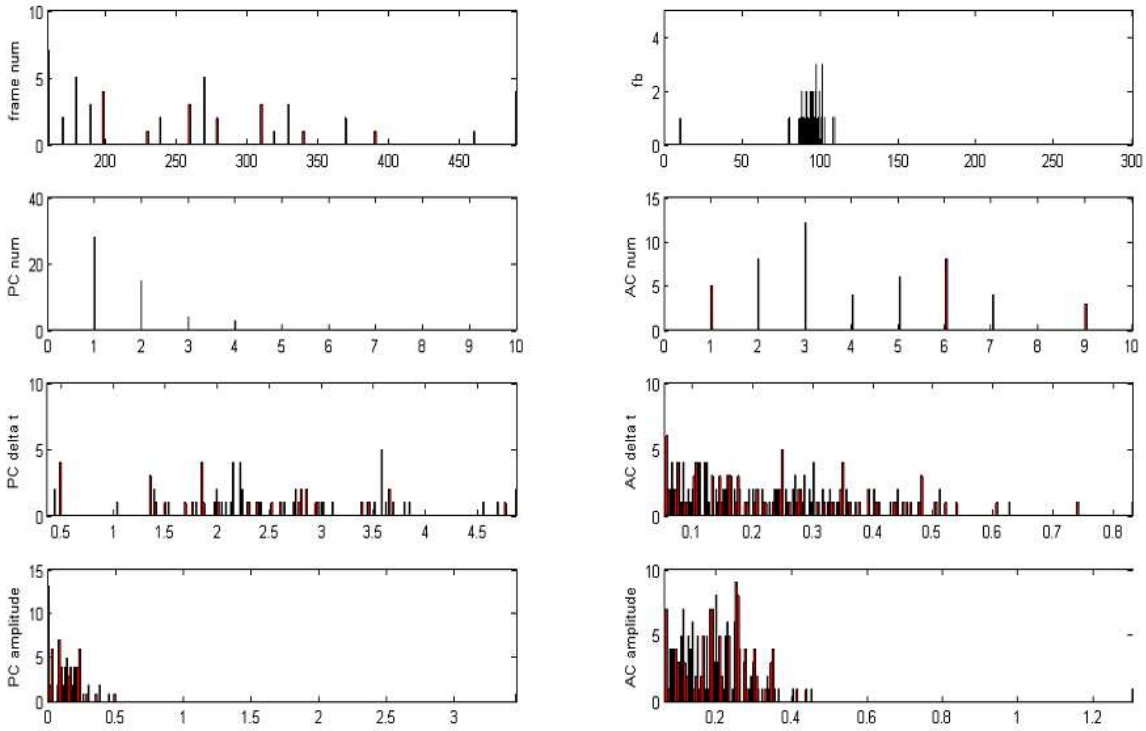

Fig. 10: Comparison of 7-parameter distributions of standard Thai train noise corrupted speech at 20 dB

DISCUSSION

Fig. 2-10 present the frequency distribution graphs. They show that the distributions of speech corrupted by four types of noises are significantly different. The distribution ranges of most parameters from speech corrupted by noises at 0 dB are mostly lower than that of 20 dB. It can be seen that the air-conditioner and car noises cause the distributions of baseline frequency split into two sub-groups. As for the parameter of number of phrase commands, there is a very little change for its distribution caused by noises.

CONCLUSION

This study presents the effects of noises on modeling of F0 contour for standard Thai. The Fujisaki's model was chosen in this study. In the experiments, four types of environmental noises are recorded with different levels of power. The model parameters have been explained and summarized. The experimental results indicate that some types of noises cause some differences in the distributions of the model parameters. All in all, the environmental 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 (SSW6), ISCA, 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, Aug. 27-31, 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

Chomphan, S., 2010c. 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., 2011a. 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

Chomphan, S., 2011b. Speech compression for noisecorrupted thai expressive speech. J. Comput. Sci., 7: 1565-1573. DOI: 10.3844/jcssp.2011.1565.1573

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.

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. 12-18, 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), 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 Xplore Press, Beijing, China, 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, ISCA Archive, 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 International 7th International Conference on Spoken Language Processing, Sept. 16-20, ISCA Archive, 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, May13-17, 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 International Conference on Acoustics, Speech and Signal Processing, Apr. 6-10, IEEE Xplore Press, Hong Kong, 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.

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. Proceedings of the International Symposium on Tonal Aspects of Languages, (ISTAL' 06), La Rochelle, France.