American Journal of Applied Sciences, 10 (2): 185-188, 2013
ISSN: 1546-9239

doi:10.3844/ajassp.2013.185.188 Published Online 10 (2) 2013 (http://www.thescipub.com/ajas.toc)

Thai Dialects and Structural Modeling with Noises

[1, 2]Suphattharachai Chomphan, [1]Numchok Budwong and [1]Sirarot Preechatanapoj

[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 2012-12-28, Revised 2013-03-10; Accepted 2013-04-12

ABSTRACT

Fundamental Frequency (F0) conveys the prosodic information of the human speech. The modeling of the dialects' F0 in a particular language is vital issue that should be taken into account. Four main dialects are spoken in different regions of Thailand including central, north, northeast and south regions. Another important issue is the environmental noises which is often be perceived in the daily life and causing the degradation in speech quality. The robustness of the F0 modeling techniques can be evaluated by studying the effects of noises for Thai dialects. The structural model has been chosen in this study. The four-type background noises with five different levels of power are applied in this study. The synthesized F0 from the structural model has been compared with the F0 from natural speech with different scenarios including noise types, noise levels speech dialects and speech genders. From the experimental results, the root mean square errors between the synthesized F0 and the natural F0 are calculated. When increasing the noise level, the root mean square error decreases. As for the different noise types, air-conditioner noise gives the highest level of root mean square error, while the train noise brings the lowest level of root mean square error. As for the different male speech dialects, center and northeast dialects are rather higher than those of north and south dialects. As for the different female speech dialects, north dialect has the smallest deviation among all dialects. As for the different genders, female speech give higher root mean square error than male speech for all types of noises and all power levels of noises. By using the structural model, the results confirm that all Thai dialects response the proposed model differently. Moreover, all four types of simulated noises deteriorate the F0 contours of all dialects differently.

Keywords: Fundamental Frequency Modeling, Environmental Noise, Fundamental Frequency, Structural Model, Speech Dialects, Speech Genders

1. INTRODUCTION

In the noisy environment, the effects of noises in various types are needed to be taken into account for human speech communication. The modeling of speech F0 contour with noises causes the degradation in intelligibility and naturalness of the speech (Chomphan and Kobayashi, 2007a). It is important to study how the noise degrades the model parameters and the resulted F0 in the modern speech processing systems. Structural modeling of fundamental frequency for Thai tones conducted in 2012 shows the effectiveness for a limited-domain speech tone corpus (Chomphan, 2012). It can be seen that the structural model parameters can be used to model Thai tones appropriately.

Moreover, Fujisaki's modeling of F0 contours for Thai Dialects has been conducted by (Chomphan, 2010b). In has been noted that Thai dialect speech corpus

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

and the effects of noises have not been studied for structural modeling (Chomphan and Kobayashi, 2007b). In this study, we applied the structural model with Thai dialects and various types of noises. The study proposes an analysis of F0 modeling of sl model in the preliminary studies has been chosen in this study. The different types of noises including air-conditioner, car, factory and train noises, are recorded and used in this study.

2. MATERIALS AND METHODS

2.1. Structural Model

This modeling technique is basically modified from mechanical vibration system. It is applied to model the F0 contour of human speech in a logarithmic scale, as depicted in **Fig. 1**. In Ni and Hirose (2006) exploited the mathematical equations to control the structure of speech F0 contour consisting of placing a series of normalized F0 targets along the time axis, which are also specified by transition time and amplitudes. It has been noted that the transitions between targets are approximated by connecting truncated second-order transition functions (Ni and Hirose, 2006; Seresangtakul and Takara, 2003). Fujisaki stated that F0 trajectory moves linearly with vocal-fold elongation x (Fujisaki and Sudo, 1971; Mixdorff and Fujisaki, 1997; Chomphan and Kobayashi, 2008), as shown in Equation 1:

$$\ln f_0 = \frac{b}{2}x + \ln(\sqrt{a}c_0) \quad (1)$$

where, a, b and c_0 are constant coefficients.

2.2. Experimental Design

The procedure of the experiment is presented in **Fig. 2**. First, constructing the dialect database including central, north, northeast and south dialects has been inaugurated. There is male and female speech in the corpus. Meanwhile, the noise database has been built with four different types including air-conditioner, car, factory and train noises. Thereafter, the F0 contours of clean dialect speech has been extracted in the "calculation of F0 contour" stage. Simultaneously, the clean speech from dialect speech database has been mixed with all four types of noises from the noise database in the "noises mixing noises with clean speech" stage. Subsequently, the F0 contours of noise-corrupted speech have been calculated in another stage of "calculation of F0 contour". The differences in terms of RMSE have been thereafter extracted in the "RMSE calculation" stage (Chomphan, 2010a). Finally, the RMSE values have been analyzed in the "data analysis" stage.

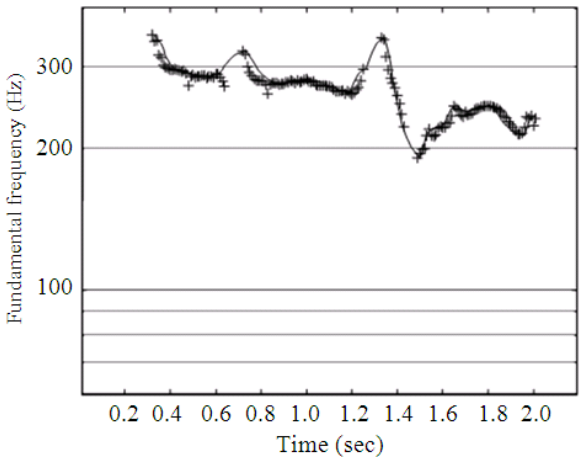

Fig. 1. An example of Thai speech F0 contour with a trend line in a logarithmic scale

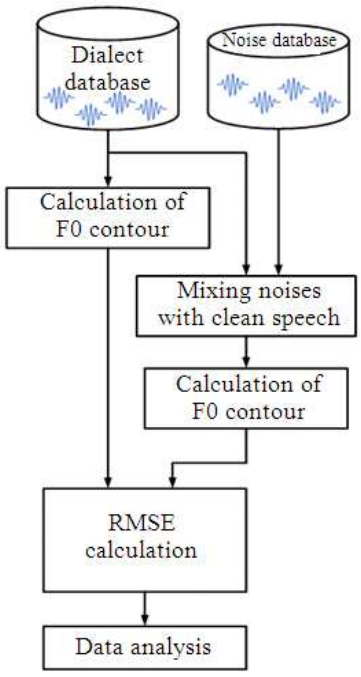

Fig. 2. Designed procedure for structural modeling

2.3. Environmental Noises

Four types of noises including air-conditioner, car, factory and train have been chosen (Chomphan and Kobayashi, 2009). They are mixed directly with the pre-recorded clean dialect speech. Before mixing noises with the clean speech, the noise magnitudes have been adjusted in five levels varying from 0 to 20 dB, with an increasing step interval of 5 dB.

3. RESULTS

From our dialect database, ten sentences in Thai for male and female genders are exploited. These sentences have been recorded in four Thai dialects of standard Thai (Center-dialect), Lanna dialect (North-dialect), Lao-style dialect (Northeast-dialect) and south regional dialect (South-dialect). Each dialect contains one hundred utterances of samples.

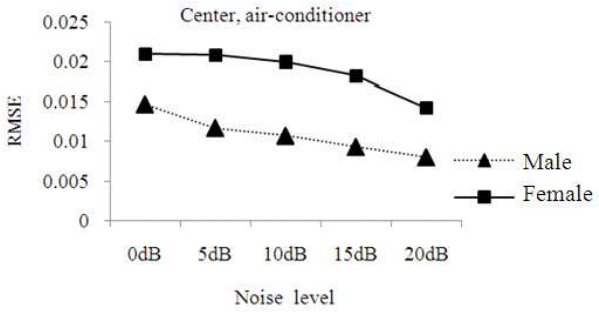

Fig. 3. RMSEs for Center dialect with or both genders at various levels of noise ranging from 0dB to 20dB

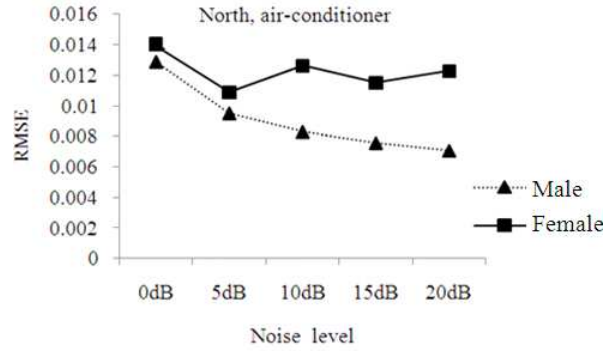

Fig. 4. RMSEs for North dialect with or both genders at various levels of noise ranging from 0dB to 20dB

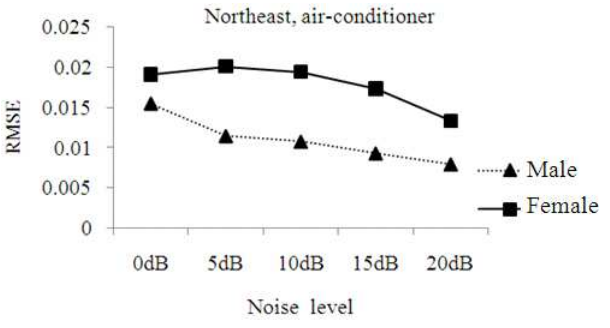

Fig. 5. RMSEs for Northeast dialect with or both genders at various levels of noise ranging from 0dB to 20dB

In the "data analysis" stage of the designed procedure for structural modeling, two aspects of analysis have been performed. The RMSEs of the structural model with all types of noises and all dialects are calculated in various levels of noise ranging from 0dB to 20dB as shown in **Fig. 3-6**. Another aspect, the RMSEs for all types of noises are compared by varying dialects as depicted in **Fig. 7-8**. Moreover, the comparison between genders is conducted for both aspects.

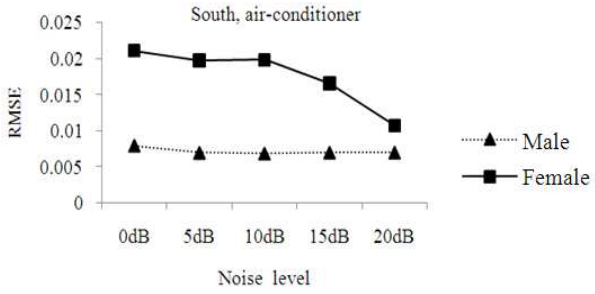

Fig. 6. RMSEs for South dialect with or both genders at various levels of noise ranging from 0dB to 20dB

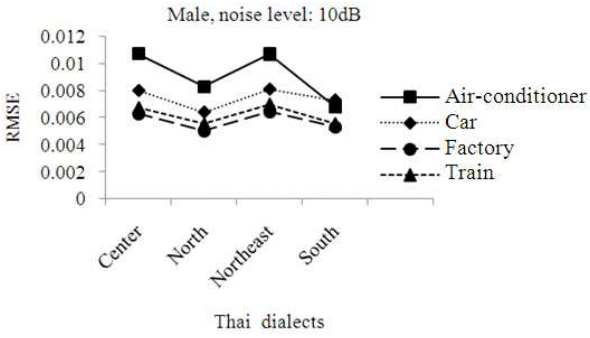

Fig. 7. RMSEs of male speech with the level of noise at 10dB for all types of noises among all dialects

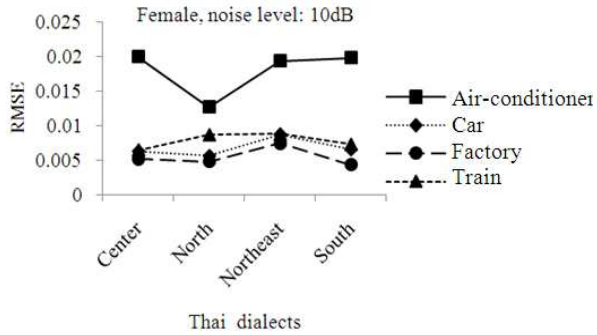

Fig. 8. RMSEs of female speech with the level of noise at 10dB for all types of noises among all dialects

4. DISCUSSION

From the experimental results, the root mean square errors between the synthesized F0 reconstructed from structural model and the natural F0 are calculated. Most results from **Fig. 3-6** show that when increasing the noise level, the root mean square error decreases. As for the different noise types from **Fig. 7-8**, air-conditioner noise gives the highest level of root mean square error, while the train noise brings the lowest level of root mean square error. Moreover, as for the different male speech dialects, center and northeast dialects are rather higher than those of north and south dialects. It has been seen that, as for the different female speech dialects, north dialect has the smallest deviation among all dialects. Last but not least, as for the different genders, female speech give higher root mean square error than male speech for all types of noises and all power levels of noises.

5. CONCLUSION

This study presents a study of effects of noises on structural modeling of F0 contour for Thai dialects. It has been noticed that when increasing the noise level, the root mean square error decreases. As for the different noise types, air-conditioner noise gives the highest level of root mean square error, while the train noise brings the lowest level of root mean square error. As for the different male speech dialects, center and northeast dialects are rather higher than those of north and south dialects. As for the different female speech dialects, north dialect has the smallest deviation among all dialects. As for the different genders, female speech give higher root mean square error than male speech for all types of noises and all power levels of noises. The results confirm that all Thai dialects response the proposed model differently. Moreover, all four types of simulated noises deteriorate the F0 contours of all dialects differently.

6. ACKNOWLEDGEMENT

The researcher is grateful to Kasetsart University for the research grants through the Kasetsart University Research and Development Institute and 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. 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, 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., 2012. Structural modeling of fundamental frequency contour for Thai tones. Am. J. Applied Sci., 9: 1736-1741. DOI: 10.3844/ajassp.2012.1736.1741

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 Euro Speech, 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

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, pp: 452-455. DOI: 10.1109/ICASSP.2003.1198815