American Journal of Applied Sciences 9 (8): 1237-1241, 2012
ISSN 1546-9239

Effects of Environmental Noises on Fundamental Frequency Contours of Thai Expressive Speech

[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: The expressive speech of Thai had been studied for a short period of time. An important feature of speech was fundamental frequency (F0) which defines the human speech prosody. It could be used to distinguish the difference between several types of expressive speech. The environmental noises affect the F0 contour for Thai dialects as concluded in the previous study. The study prosodic information of Thai speech with various speaking styles and several types of noises had not been conducted. **Approach:** Four different types of speaking styles were used; meanwhile four types of environmental noises were recorded with different levels of power. They were subsequently mixed together. The F0 contours from different types of speaking styles, different types of noises and different levels of noises were extracted. The Root Mean Square Error (RMSE) between the F0 contour of clean speech and the noise-corrupted speech was calculated. **Results:** In the experiments, four types of noises were included train, factory, car and air conditioner. Each type of speaking style included 10 samples of 10 utterances of male and female speech. Five levels of noises were varied from 0-20 dB compared with the clean speech. It could be notified that the effects of distinguishing types of noises were different. Four different types of speaking styles were also caused the differences in RMSEs. **Conclusion:** The recorded noises deteriorate the F0 contours for all types of speaking styles in Thai.

Key words: Root Mean Square Error (RMSE), among several types, recorded noises deteriorate, simulated noises deteriorate, speech database

INTRODUCTION

In human speech production, fundamental frequency or F0 is a very crucial feature known to carry prosodic information. The intelligibility and the naturalness of speech are considerably determined by this frequency. Most of speech processing technologies must take into account this feature. In the recent study on modeling of F0 contour with noisy environment, the simulated noises deteriorate the Fujisaki's model parameters (Fujisaki and Sudo, 1971; Mixdorff and Fujisaki, 1997; Seresangtakul and Takara, 2003). However the study on the direct effect of noises on the fundamental frequency contour of the expressive speech has not been conducted. (Chomphan, 2010a; 2010b). This study proposes an analysis the differences between the fundamental frequency of clean expressive speech and noise-corrupted expressive speech in term of RMSE. Fig. 1 presents an example of fundamental frequency contour. This study concentrates on expressive speech of angry, sadness, enjoy and reading styles, meanwhile the selected four types of noises are air-conditioner, car, factory and train noises.

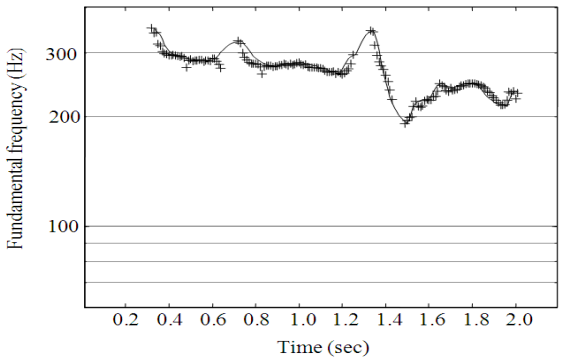

Fig. 1: An example of fundamental frequency contour

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

MATERIALS AND METHODS

Experimental design: Figure 2 shows the procedure of the experiment. The first step is to construct the expressive speech database of four styles including angry, sadness, enjoy and reading styles. Simultaneously, the noise database is constructed with four different types including air-conditioner, car, factory and train noises. Subsequently, the F0 contours of clean speech are extracted accurately in the "calculation of F0 contour" stage. Moreover, the clean speech from speech database is mixed with all four types of noises from the noise database in the "noises mixing noises with clean speech" stage.

F0 contour". The differences in terms of RMSE are then calculated in the "RMSE calculation" stage (Chomphan, 2011a). In the last stage of data analysis, RMSE values are analyzed comparatively.

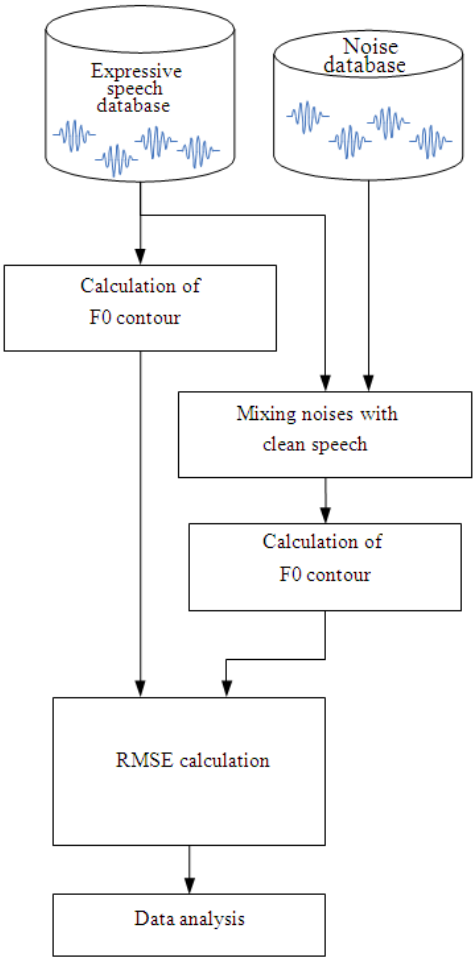

Fig. 2: The flow chart indicating the procedures in the experiment

Thereafter the F0 contours of noise-corrupted speech are extracted in another stage of "calculation of

Environmental noises: Four types of noises include train, factory, car and air conditioner. 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 five levels. These levels for each type of noise are varied from 0, 5, 10, 15, 20 dB, respectively.

RESULTS

As for the expressive speech corpus, ten sentences in Thai for both female and male speech are exploited. The sentences cover four speaking styles (reading, angry, enjoy and sadness) (Chomphan and Kobayashi, 2007a; 2007b). It has been defined that one style of speaking covers one hundred sentences. Consequently, for each gender, the corpus has four hundred sentences. (Mixdorff and Fujisaki, 1997).

From the "data analysis" stage in Fig. 2, the following charts are summarized (Chomphan, 2011b). First, the noise effects on the male-angry-style speech are summarized in terms of RMSE values with four different types of noises and five different levels of noises in Fig. 3 (Chomphan and Kobayashi, 2009; 2008). Second, the noise effects on the male-sad-style speech are summarized in terms of RMSE values with four different types of noises and five different levels of noises in Fig. 4. Third, the noise effects on the male-enjoy-style speech are summarized in terms of RMSE values with four different types of noises and five different levels of noises in Fig. 5. Fourth, the noise effects on the male-reading-style speech are summarized in terms of RMSE values with four different types of noises and five different levels of noises in Fig. 6. Fifth, the noise effects on the female-angry-style speech are summarized in terms of RMSE values with four different types of noises and five different levels of noises in Fig. 7. Sixth, the noise effects on the female-sad-style speech are summarized in terms of RMSE values with four different types of noises and five different levels of noises in Fig. 8. Seventh, the noise effects on the female-enjoy-style speech are summarized in terms of RMSE values with four different types of noises and five different levels of noises in Fig. 9. Eighth, the noise effects on the female-read-style speech are summarized in terms of RMSE values with four different types of noises and five different levels of noises in Fig. 10.

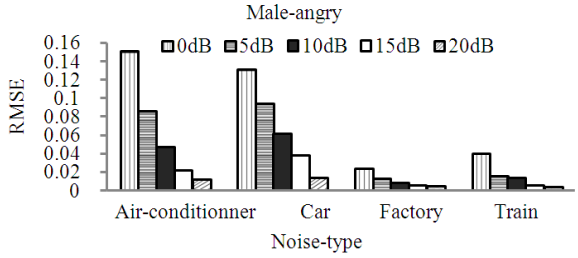

Fig. 3: Noise effects in RMSEs on the male-angry-style speech for four different types of noises and five different levels of noises

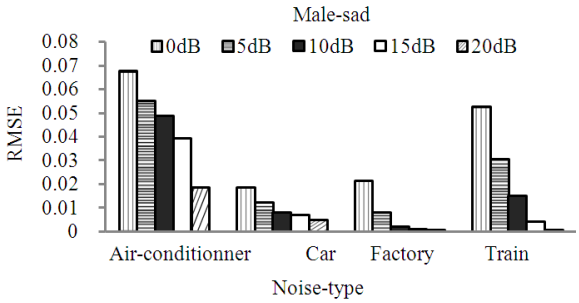

Fig. 4: Noise effects in RMSEs on the male-sad-style speech for four different types of noises and five different levels of noises

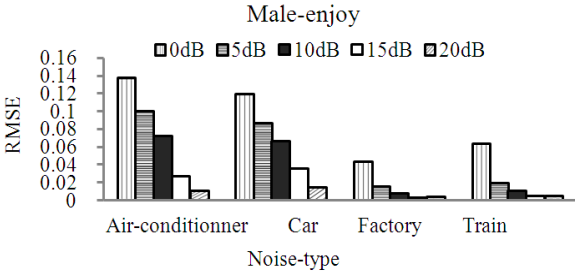

Fig. 5: Noise effects in RMSEs on the male-enjoy-style speech for four different types of noises and five different levels of noises

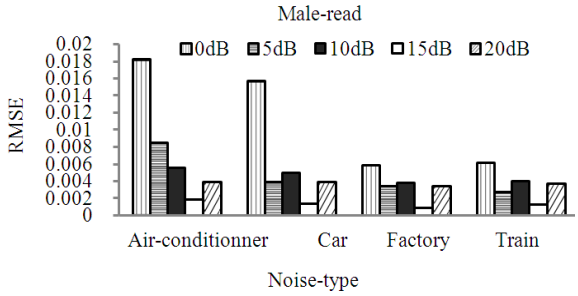

Fig. 6: Noise effects in RMSEs on the male-reading-style speech for four different types of noises and five different levels of noises

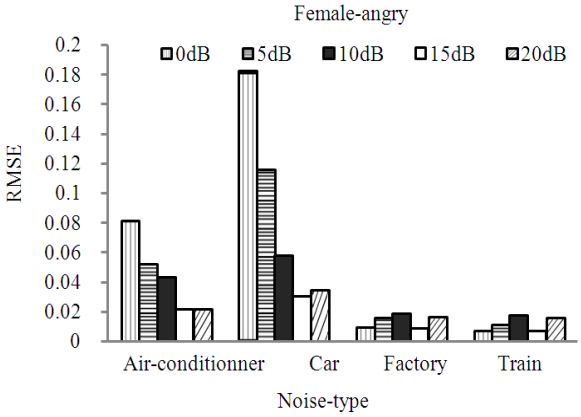

Fig. 7: Noise effects in RMSEs on the female-angry-style speech for four different types of noises and five different levels of noises

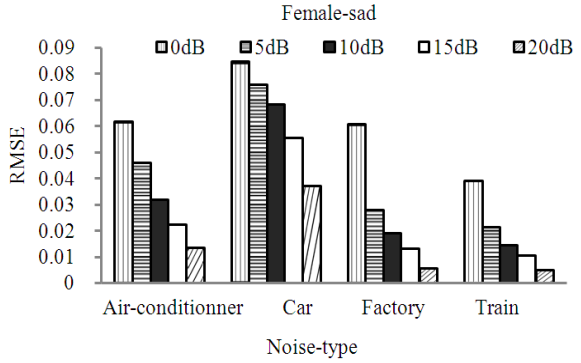

Fig. 8: Noise effects in RMSEs on the female-sad-style speech for four different types of noises and five different levels of noises

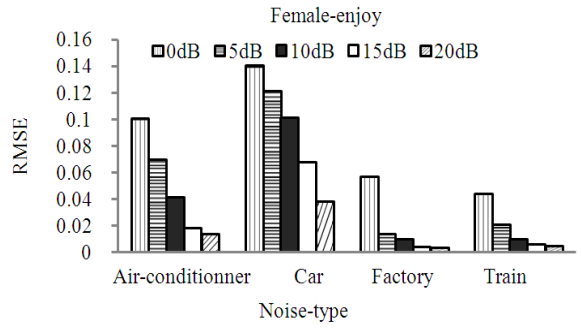

Fig. 9: Noise effects in RMSEs on the female-enjoy-style speech for four different types of noises and five different levels of noises

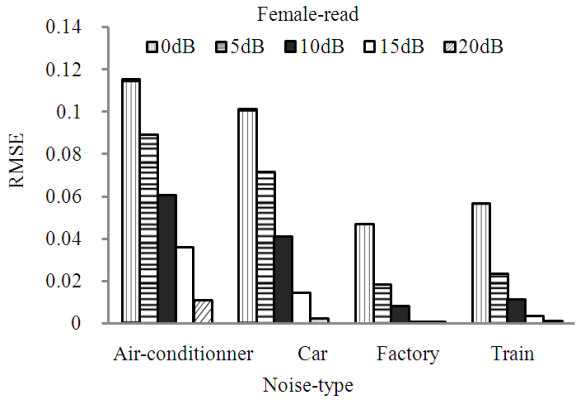

Fig. 10: Noise effects in RMSEs on the female-reading-style speech for four different types of noises and five different levels of noises

DISCUSSION

From Fig. 3-6 showing the noise effects on male speech, the reading-style gives the lowest level of RMSEs, meanwhile the angry-style and enjoy-style both give the highest levels of RMSEs. It can be inferred that the reading-style speech is the most robustness among all speaking styles. When comparing among noise types, it can be seen that male speech gives the higher RMSEs than that of male speech for all four types of noises. Then comparing among types of noises, the effect of air-conditioner noise is mostly highest for all speaking styles. Moreover the effect of factory noise is mostly lowest.

When considering the level of noises, the order of RMSEs are as follows; 0dB, 5dB, 10dB, 15dB and 20dB, respectively. The 0dB-level means the power of clean speech and the power of noise are equal, therefore the effect of noise is highest. Meanwhile, the 20dB-level means the power of clean speech is approximately 100 times higher than the power of noise; therefore the effect of noise is lowest. As for the noise effects on female speech from Figs. 7-10, it can be seen that all speaking styles do not show significantly different from each others. When comparing among noise types, it can be noticed that the car noise gives mostly highest level of RMSEs except for that of the reading-style speech. Meanwhile, the factory and train noises give lowest level of RMSEs. When considering the level of noises, the result corresponds to that of male speech.

CONCLUSION

This study presents a study of effects of noises on fundamental frequency contour for Thai expressive speech. Four types of environmental noises are recorded with five different levels of power. The differences of fundamental frequency contours between the noise-corrupted samples and the clean samples are calculated in terms of RMSEs. The simulated noises deteriorate fundamental frequency contours differently depending on type of noise, level of noise, gender and speaking style.

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. Proceedings 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. Tokyo Institute of Technology. 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., 2011a. Analysis of fundamental frequency contour of coded speech based on multipulse based code excited linear prediction algorithm. J. Comput. Sci., 7: 865-870. DOI: 10.3844/jcssp.2011.865.870

Chomphan, S., 2011b. 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. 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. Proceedings of the 5th European Conference on Speech Communication and Technology, Sept. 22-25, ISCA Archive, Rhodes, Greece, pp: 187-190.

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