Journal of Computer Science 7 (4): 568-572, 2011
ISSN 1549-3636

Effects of Noises on the Analysis of Fundamental Frequency Contours for Thai Speech

Suphattharachai Chomphan
Department of Electrical Engineering, Faculty of Engineering at Si Racha,
Kasetsart University, 199 M.6, Tungsukhla, Si Racha, Chonburi, 20230, Thailand

Abstract: Problem statement: In speech communication, noises from surrounding environments affect the communication quality with various aspects. The received speech quality should be analyzed to see how the important noises reduce the speech quality so that we can eliminate them in an appropriate way. **Approach:** This study presents a study on the analysis of the noise effects on Thai speech. Four kinds of noises; air conditioner, car, factory and train, are chosen to be simulated in the study. The various levels of signal-to-noise ratios are conducted. The root mean square error between the fundamental frequency contours of the corrupted speech and the clean speech is calculated. Finally, the analysis of the root mean square error in terms of comparisons among genders, the four kinds of noises and various levels of signal-to-noise ratios is performed. **Results:** In the experiments, 400 speech utterances of male and female are used as speech materials. The average values of root mean square error are calculated. The results show that the fundamental frequency contour of female speech is affected more than that of male speech. Comparing among four kinds of noises, the car noise has the highest influence, while the factory noise has the lowest influence. Moreover, the root mean square error is inversely proportional to the level of signal-to-noise ratio. **Conclusion:** From the finding, the noises from surrounding environments have affected the speech quality of fundamental frequency contour. This study is the preliminary knowledge to enhance the speech quality for further works such as speech synthesis systems or other speech processing technologies.

Key words: Noised speech, speech analysis, fundamental frequency contour, speech enhancement, root mean square error, surrounding environments, signal-to-noise ratios, data material, noise-merged signal, extracted fundamental

INTRODUCTION

In speech technology study, speech analysis has been conducted for many languages. It is a preliminary procedure in the speech processing area; including speech recognition, speech synthesis, speech analysis and speech coding (Chomphan and Kobayashi, 2007; Chomphan and Kobayashi, 2008; Chomphan and Kobayashi, 2009). In practical situation, noises from surrounding environments affect the speech quality with various aspects. The fundamental frequency extracted frame-by-frame from the speech is an important feature indicating the pitch or voicing level of the speech. It has been widely exploited in most of speech processing technology mentioned above. The study of the affect of surrounding noises to the fundamental frequency should be conducted appropriately. The important background noises include car noise, train noise, factory noise and air conditioner noise (Manohar and Rao, 2006). The study concentrates how the noises affect the fundamental frequency contour of the speech by varying the level of signal-to-noise ratio. It is expected to apply the finding knowledge in further study in advanced research such as speech synthesis and recognition (Chomphan, 2009; Chomphan, 2010a; Chomphan, 2010b; Chomphan, 2010c).

MATERIALS AND METHODS

Fundamental Frequency Contour (F0 contour): There is a substantial amount of data on the frequency of the voice fundamental or fundamental frequency (F0) in the speech of speakers who differ in age and sex. The data have been published for several languages and for various types of discourse. The data always include an average measure of F0, usually expressed in Hz, but in some cases the average duration of a period has been reported instead. Typical values obtained for F0 are 120 Hz for male speech and 210 Hz for female speech (Waldstein and Boothroyd, 1994). An example of F0 contour of the natural speech is depicted in Fig. 1.

Typically, the mean values of F0 change slightly with age. For female speech, F0 is quite stationary up to the period of menopause, when it decreases to reach the minimum which is about 15 Hz lower around 70 years of age (Pegoraro-Krook, 1988). The physiological changes is an effect of the increased testosterone-oestrogen ratio at that period. A similar decreasing of F0 can be caused by the habit of smoking (Gilbert and Weismer, 1974). For male speech, the dramatic decrease in F0 during puberty duration has been observed to continue with subsequent deceleration until about 35 years of age. Thereafter, at about 55 years of age, F0 begins to rise again (Pegoraro-Krook, 1988).

F0 modeling is another issue that is related to this study. The former study on F0 modeling has been widely performed in various speech units and several techniques such as utterance level (Fujisaki and Ohno, 1998; Fujisaki *et al*., 1990; Tao *et al*., 2006; Saito and Sakamoto, 2002 Ni and Hirose, 2006; Li *et al*., 2004), word and syllable levels (Fujisaki *et al*., 1990). In Thai speech, Fujisaki's model has been successfully applied for modeling of utterances, tones and words (Hiroya and Sumio, 2002; Seresangtakul and Takara, 2002; Seresangtakul and Takara, 2003).

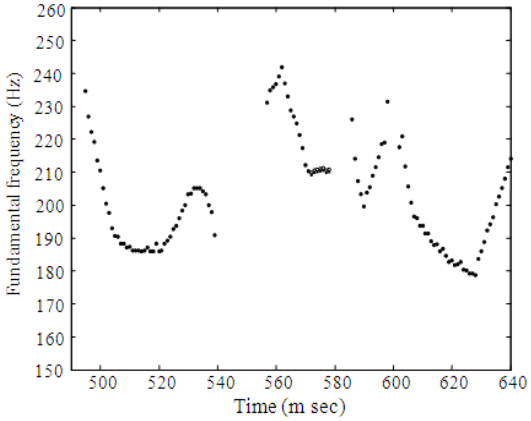

Fig. 1: An example of fundamental frequency contour of female speech

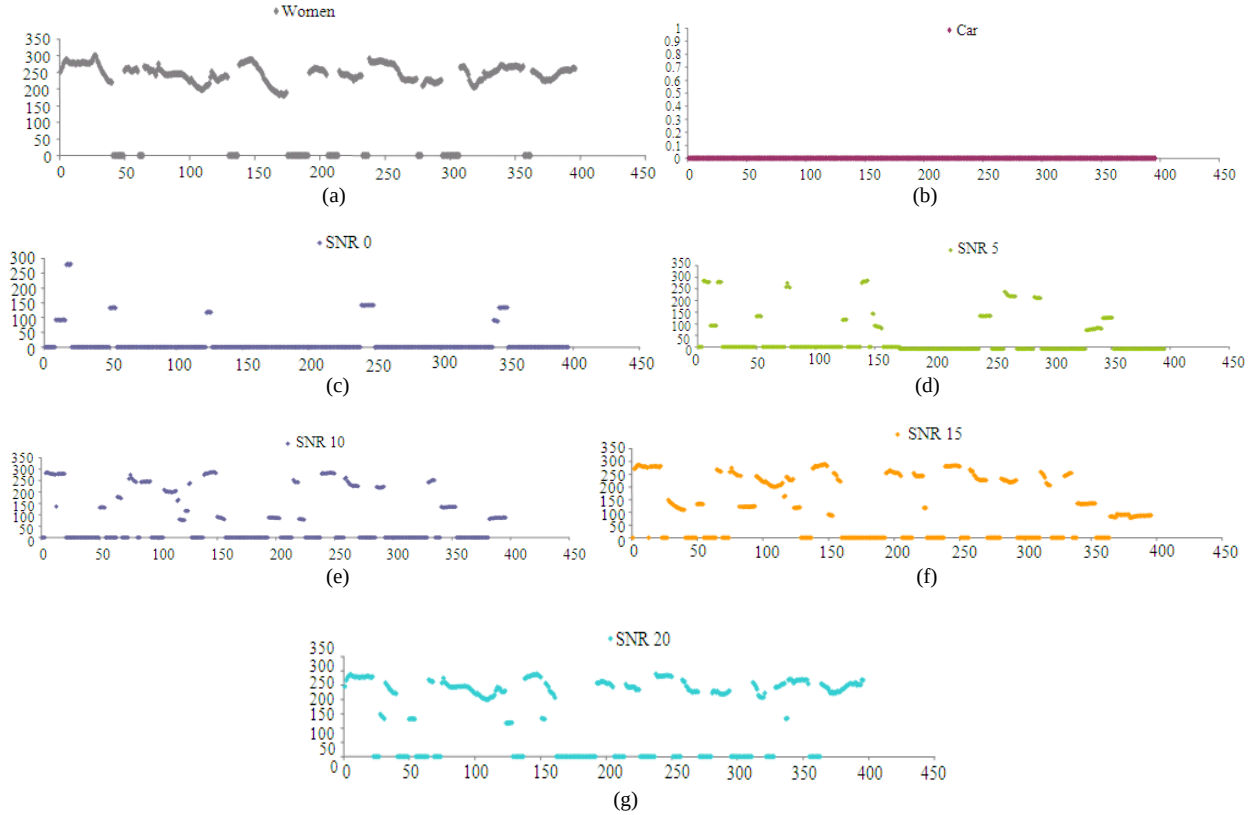

Fig. 2: Examples of fundamental frequency contours of female speech extracted from (a) clean speech signal, (b) pure car noise signal, (c) car noise-merged signal with SNR 0dB, (d) car noise-merged signal with SNR 5dB, (e) car noise-merged signal with SNR 10dB, (f) car noise-merged signal with SNR 15dB and (g) car noise-merged signal with SNR 20dB

Types of noise: The interesting background noises include car noise, train noise, factory noise and air conditioner noise. Each type of the noises is recorded separately and its amplitude is scaled to acquire the desired level of noise. In other words, the energy of noise and the speech signal are calculated then these signals are merged with five levels of SNRs of 0, 5, 10, 15 and 20 dB (Shareha *et al*., 2009; Geravanchizadeh and Rezaii, 2009; Rushaidin *et al*., 2009; Rajarathinam and Parmar, 2011).

Procedures of fundamental frequency contour analysis: The following analysis procedures are implemented for an utterance from the speech data material and noise data material (Abdellaoui, 2009; Chomphan, 2010b; 2010c; 2010d; 2010e; Lampson *et al*., 2010; Ramadan, 2010; Teymourzadeh *et al*., 2010):

- Scaling of noise signal to obtain five desired level comparing with clean speech signal
- Merging noise signal with the clean speech signal
- Extracting F0 contour from both the noise-merged signal and the clean speech signal
- Calculating the Root Mean Square Error (RMSE) between the F0 contours of noise-merged signal and the clean speech signal
- Calculating the statistical values of the root mean square error in step 4

It has been noted that these analysis procedures are conducted for all types of noises. Figure 2 shows the examples of fundamental frequency contours of female speech extracted from speech signal with different situations including clean speech signal, pure car noise signal, car noise-merged signal with SNR 0dB, car noise-merged signal with SNR 5dB, car noise-merged signal with SNR 10dB, car noise-merged signal with SNR 15dB and car noise-merged signal with SNR 20dB. It can be seen that the higher level of noise (or lower level of SNR) can deteriorate the F0 contour from the original clean speech. Moreover, no F0 value can be extracted from pure noise signal as seen in Fig. 2b.

RESULTS

By using the speech database of 200 sentences of female speech and 200 sentences of male speech, the extracted fundamental frequency contours of noise-merged signal with SNR 0dB, noise-merged signal with SNR 5dB, noise-merged signal with SNR 10dB, noise-merged signal with SNR 15dB and noise-merged signal with SNR 20dB and the clean speech signal are extracted. Root mean square error between the extracted fundamental frequency values of the noise-merged signal and the clean speech signal are calculated for all sentences in the speech database. Four important types of noises which are simulated in this study include air conditioner noise, car noise, factory noise and train noise. Figure 3-6 show the root mean square errors for different levels of SNRs, meanwhile the comparison of RMSEs between the female speech and the male speech is present for all figures.

DISCUSSION

From the comparison of root mean square errors between the female corrupted speech and the corrupted male speech for different levels of SNRs in Fig. 3-6, it can be seen that the RMSEs of female speech are mostly higher than those of male speech. Comparing among four kinds of noises in Fig. 3-6, the car noise in Fig. 4 has the highest influence, while the factory noise in Fig. 5 has the lowest influence averagely. The RMSE decreases when the SNR is

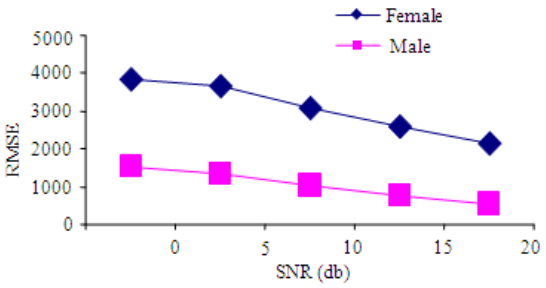

Fig. 3: Comparison of root mean square errors between the female corrupted speech and the corrupted male speech for different levels of SNRs with air conditioner noise

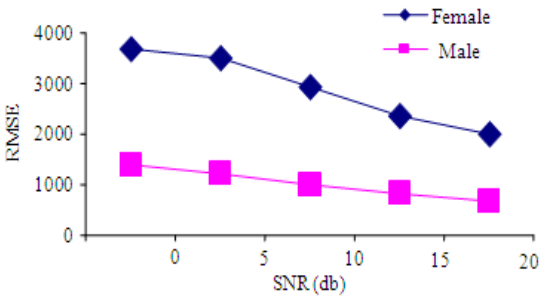

Fig. 4: Comparison of root mean square errors between the female corrupted speech and the corrupted male speech for different levels of SNRs with car noise

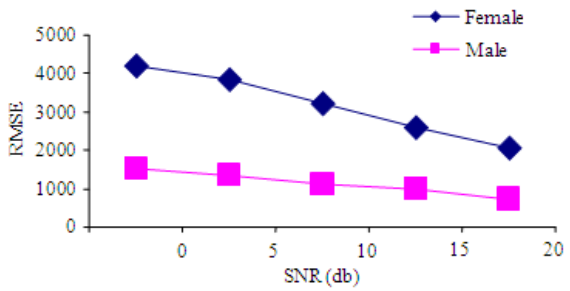

Fig. 5: Comparison of root mean square errors between the female corrupted speech and the corrupted male speech for different levels of SNRs with factory noise

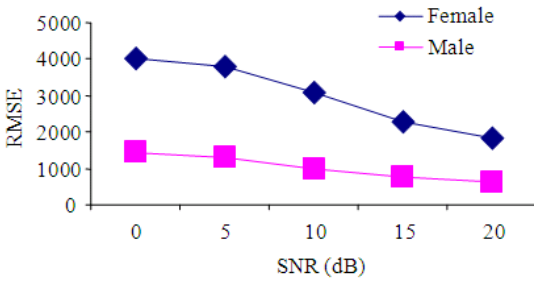

Fig. 6: Comparison of root mean square errors between the female corrupted speech and the corrupted male speech for different levels of SNRs with train noise

increasing for both male and female speech and for all types of noises. It is the result from the less effect of noise. All in all, the root mean square error is inversely proportional to the level of signal-to-noise ratio.

CONCLUSION

This study proposes a study on the analysis of the noise effects on Thai speech. Four interesting kinds of noises; air conditioner, car, factory and train, are mainly focused. The various levels of signal-to-noise ratios are performed. The root mean square error between the fundamental frequency contours of the corrupted speech and the clean speech is calculated. Finally, the analysis of the root mean square error in terms of comparisons among genders, the four kinds of noises and various levels of signal-to-noise ratios is conducted. We use 400 of male and female utterances in the study. The results show that the fundamental frequency contour of female speech is affected more than that of male speech. Comparing among four kinds of noises, the car noise has the highest influence, while the factory noise has the lowest influence. Moreover, the root mean square error is inversely proportional to the level of signal-to-noise ratio. All in all, the noises from surrounding environments have affected the speech quality of fundamental frequency contour.

ACKNOWLEDGEMENT

The author is grateful to J. Nargwi-graikit and P. Paholthap for providing the speech database.

REFERENCES

Abdellaoui, M., 2009. Inverse sine phase detector phase locked loop associated with modified multi band lc quadrature voltage controlled oscillator for wireless communication systems at 0.9, 1.8, 2.4, 3.5 GHz. Am. J. Eng. Applied Sci., 2: 328-336. DOI: 10.3844/ajeassp.2009.328.336

Chomphan, S. and T. Kobayashi, 2007. 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, Tokyo Institute of Technology, Japan, 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., 2009. Towards the development of speaker-dependent and speaker-independent hidden markov model-based thai speech synthesis. J. Comput. Sci., 5: 905-914. DOI: 10.3844/jcssp.2009.905.914

Chomphan, S., 2010a. Fujisaki's model of fundamental frequency contours for thai dialects. J. Comput. Sci., 6: 1246-1254. DOI: 10.3844/jcssp.2010.1246.1254

Chomphan, S., 2010b. Multi-pulse based code excited linear predictive speech coder with fine granularity scalability for tonal language. J. Comput. Sci., 6: 1288-1292. DOI: 10.3844/jcssp.2010.1288.1292

Chomphan, S., 2010c. Performance evaluation of multi-pulse based code excited linear predictive speech coder with bitrate scalable tool over additive white gaussian noise and rayleigh fading channels. J. Comput. Sci., 6: 1433-1437. DOI: 10.3844/jcssp.2010.1433.1437

Chomphan, S., 2010d. 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., 2010e. Tone question of tree based context clustering for hidden markov model based thai speech synthesis. J. Comput. Sci., 6: 1468-1472. DOI: 10.3844/jcssp.2010.1468.1472

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 Spoken Language Processing, Oct. 12-16, IEEE Xplore, Beijing, China, 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 1st International Conference on Spoken Language Processing, Nov. 18-22, University of Tokyo, Japan, pp: 841-844.

Geravanchizadeh, M. and T.Y. Rezaii, 2009. Transform domain based multi-channel noise cancellation based on adaptive decorrelation and least mean mixed-norm algorithm. J. Applied Sci., 9: 651-661.

Gilbert, H.R. and G.G. Weismer, 1974. The effects of smoking on the speaking fundamental frequency of adult women. J. Psychol. Res., 3: 225-231. DOI: 10.1007/BF01069239

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, University of Tokyo, Japan, pp: 516-519. DOI: 10.1109/ICOSP.2002.1181106

Lampson, B., Y. Han, A. Khalilian, J. Greene and R.W. Mankin *et al*., 2010. Characterization of substrate-borne vibrational signals of euschistus servus (heteroptera: pentatomidae). Am. J. Agric. Bio. Sci., 5: 32-36. DOI: 10.3844/ajabssp.2010.32.36

Li, Y., T. Lee and Y. Qian, 2004. Analysis and modeling of F0 contours for cantonese text-to-speech. ACM Trans. Asian Lang. Inform. Process., 3: 169-180. DOI: 10.1145/1037811.1037813

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

Pegoraro-Krook, M.I., 1988. Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis. Folia Phoniatrica, 40: 82-90. DOI: 10.1159/000265888

Rajarathinam, A. and R.S. Parmar, 2011. Application of parametric and nonparametric regression models for area, production and productivity trends of castor (Ricinus communis L.) crop. Asian J. Applied Sci., 4: 42-52.

Ramadan, Z., 2010. Error vector normalized adaptive algorithm applied to adaptive noise canceller and system identification. Am. J. Eng. Applied Sci., 3: 710-717. DOI: 10.3844/ajeassp.2010.710.717

Rushaidin, M.M., S.H. Salleh, T.T. Swee, J.M. Najeb and A. Arooj, 2009. Wave V detection using instantaneous energy of auditory brainstem response signal. Am. J. Applied Sci., 6: 1669-1674. DOI: 10.3844/ajassp.2009.1669.1674

Saito, T. and M. Sakamoto, 2002. Applying a hybrid intonation model to a seamless speech synthesizer. Proceedings of the 7th International Conference on Spoken Language Processing, Sept. 16-20, 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 IEEE International Conference on Acoustics, Speech, and Signal Processing, May 13-17, IEEE Xplore, USA., pp: 1-1. DOI: 10.1109/ICASSP.2002.1005787

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

Shareha, A.A.A., M. Rajeswari and D. Ramachandram, 2009. Multimodal integration (image and text) using ontology alignment. Am. J. Applied Sci., 6: 1217-1224. DOI: 10.3844/ajassp.2009.1217.1224

Tao, J., J. Yu and W. Zhang, 2006. Internal dependence based F0 model for mandarin tts system. ProceedingS of theTC-STAR Workshop on Speech-to-Speech Translation, June 19-21, Barcelona, Spain, pp: 171-174.

Teymourzadeh, R., Y.S. Algnabi, M. Othman, M.S. Islam and J.M.V. Hong, 2010. VLSI implementation of novel class of high speed pipelined digital signal processing filter for wireless receivers. Am. J. Eng. Applied Sci., 3: 663-669. DOI: 10.3844/ajeassp.2010.663.669

Waldstein, R.S. and A. Boothroyd, 1994. Speechreading enhancement using a sinusoidal substitute for voice fundamental frequency. Speech Commun, 14: 303-312. DOI: 10.1016/0167-6393(94)90024-8