American Journal of Applied Sciences 9 (10): 1736-1741, 2012
ISSN 1546-9239
© 2012 Science Publication

Structural Modeling of Fundamental Frequency contour for Thai Tones

[1, 2]Suphattharachai Chomphan
[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

Abstract: Problem statement: In Thai, tone is an essential feature of a prosodic syllable to identify the meanings of that syllable or that part of word. To generate the tonal speech with natural prosody, it is needed to manage the fundamental frequency (F0) of the speech appropriately. A successful approach of structural modeling from Mandarin Chinese has been adapted to model Thai tone. **Approach:** The structural modeling of voice F0 contours for Thai tones has been studied. Both male and female speech are concerned. The speech material covers 15 syllables with 5 tones. We use 30 samples for each syllable. The structural modeling parameters for all tones are extracted. Thereafter, the Root Mean Square (RMS) error between the re-synthesized F0 contour and the natural F0 contour is calculated. **Results:** The experimental analysis shows that RMS errors of all tones are mutually different. It has been noticed that the tone 1 or low tone has the smallest error among all tones in average. **Conclusion:** The structural model is effectively applied to model Thai tones. The structural modeling can distinguish each tone empirically.

Key words: Inherently supra-segmental, structural modeling responses, important feature among, Root Mean Square (RMS), advanced research, prosodic syllable

INTRODUCTION

In human speech production, the vocal chords vibrate at a temporal frequency to produce a semi-periodic air flow through the vocal tract. This frequency is known as the fundamental frequency of the output speech signal. It is an essential feature among other speech features which carry prosodic information of the natural speech. In the modern speech technology, e.g., speech recognition, speech analysis and synthesis, it is necessary to model the F0 with high accuracy. In the former studies, several modeling techniques have been conducted at different levels of speech units, e.g., word and syllable levels, sentence level (Tran *et al*., 2006; Tao *et al.*, 2006; Fujisaki and Ohno, 1998; Fujisaki *et al*., 1990; Saito and Sakamoto, 2002; Li *et al*., 2004; Fujisaki and Sudo, 1971). This model has been efficiently applied to Thai language in the levels of utterance, word and tone (Seresangtakul and Takara, 2002; Hiroya and Sumio, 2002; Seresangtakul and Takara, 2003). It has been noted that tone is an essential feature for a speech unit of syllable in Thai. The different tone of a syllable gives the different meanings. Modeling of tone in tonal language is very crucial in the application of speech processing.

This research presented another approach of fundamental frequency contour modeling for Thai tones. The structural model is chosen to be exploited (Ni and Hirose, 2006). The root mean square (RMS) error is calculated for evaluating the modeling quality for all tones including tone 0, tone 1, tone 2, tone 3 and tone 4 (mid tone, low tone, falling tone, high tone and rising tone).

MATERIALS AND METHODS

Structural model: The fundamental frequency contour is illustrated in Fig. 1. The mathematical model has been applied (Chomphan, 2011a; 2011b; Ni and Hirose, 2006). This contour is modeled by using a structural control which consists of locating a number of normalized fundamental freuency targets along time axis in logarithmic scale. The fundamental freuency targets or pitch targets are specified by amplitudes and transition time. This transition between any two adjacent targets is approximated by using a truncated second-order function as depicted in Fig. 2.

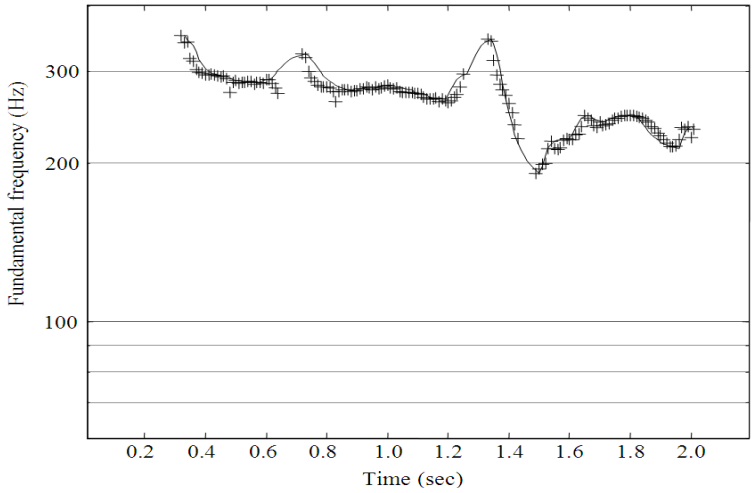

Fig. 1: F0 contour with a trend line in a logarithmic scale

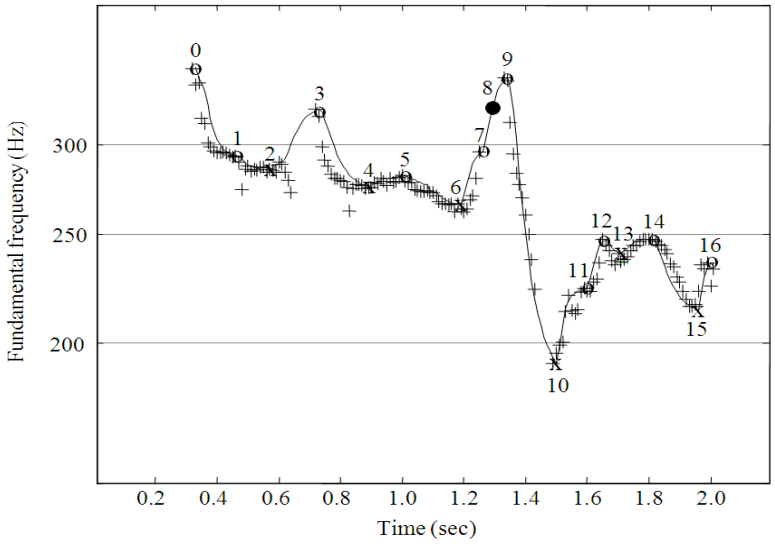

Fig. 2: Example of pitch target allocation on an F0 contour

An example of re-synthesis of fundamental frequency contour by exploitng the structural model is shown in Fig. 3. Figure 3a presents the fundamental frequency contour of the natural speech. Meanwhile Fig. 3b presents related parameter of λ with three different damping ratios ζ. Moreover, Fig. 3c illustrates re-synthesized fundamental frequency contour with different damping ratios from Fig. 3b. Finally, the natural F0 contour and the re-synthesized F0 contour are compared in Fig. 3d.

The diagram of core experiment: The Fig. 4 presents the diagram of core experiment starting from speech database to data analysis. The speech database is firstly contructed. The speech of both man and woman is recorded in syllable basis. Five Thai tones including tone 0, 1, 2, 3 and 4 are designed with the same amount and pattern. Each tone consists of 15 syllables, while each syllable consists of 30 samples. As a result, the speech database covers 4,500 speech utterances. After constructing the speech database, the fundamental frequency of an utterance are extracted. Thereafter, the pitch targets are placed by finding for all of the local minimums and maximums. An exponential function is used to approximate an appropriate route between two adjacent pitch targets.

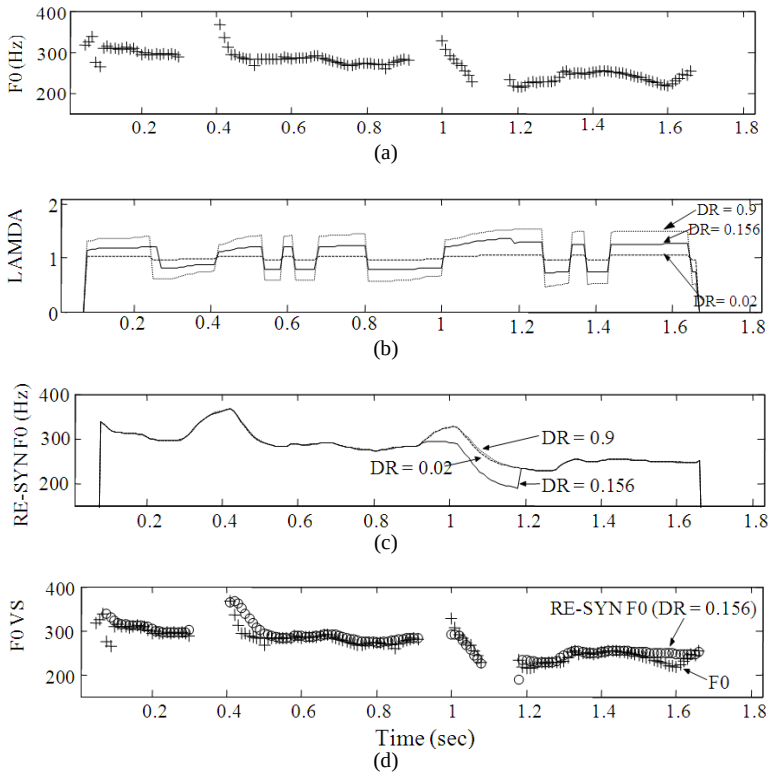

Fig. 3: An example of re-synthesis of F0 contour by using the structural model (RE-SYN denotes "resynthesized", DR denotes "damping ratio")

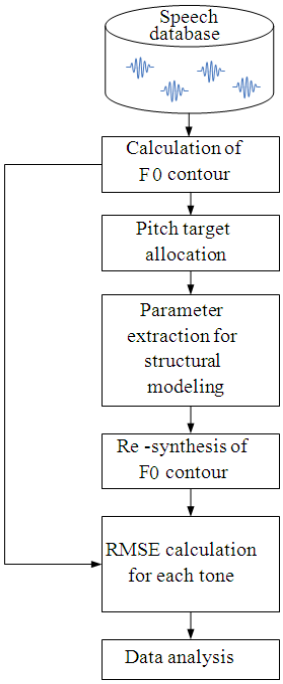

Fig. 4: Diagram of the core experiment

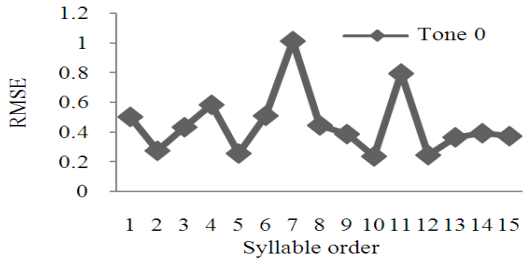

Fig. 5: Averaged RMS error for tone 0 of female speech

The extracted parameters of most of exponential functions is used to re-synthesis the fundamental frequency contour. To evaluate the difference between the natural fundamental frequency contour and the re-synthesis fundamental frequency contour, RMS error calculation is performed. Lastly, the RMS error has been statistically analyzed for all tones in Thai.

RESULTS

The evaluation of the model can be presented in the eleven figures (Fig. 5-15) resulting from the RMS error calculation process. The averaged RMS errors for all five tones of two genders have been calculated.

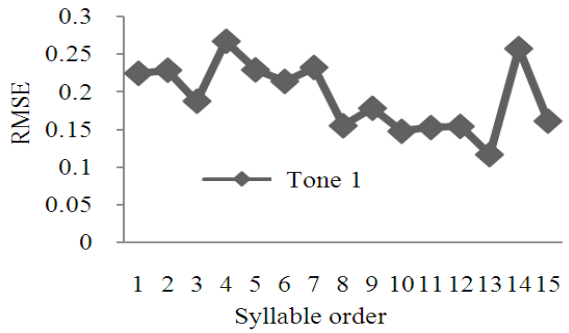

Fig. 6: Averaged RMS error for tone 1 of female speech

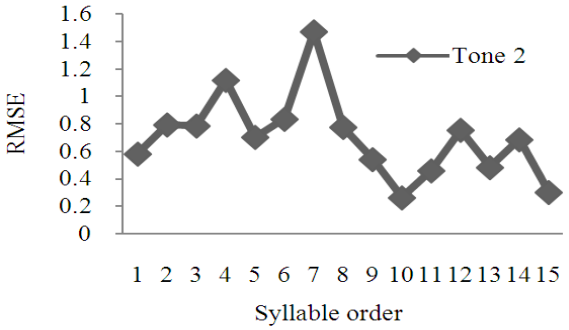

Fig. 7: Averaged RMS error for tone 2 of female speech

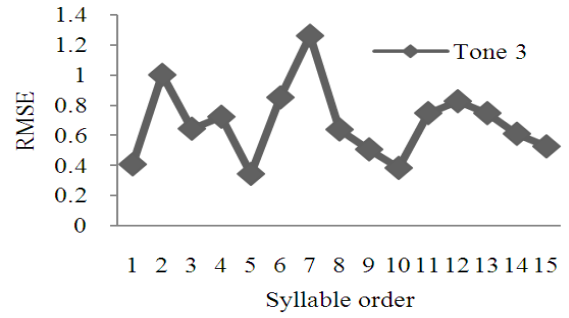

Fig. 8: Averaged RMS error for tone 3 of female speech

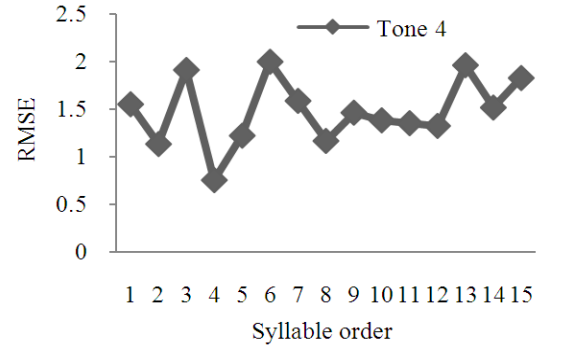

Fig. 9: Averaged RMS error for tone 4 of female speech

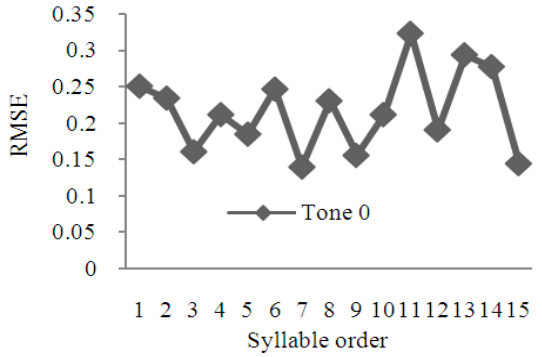

Fig. 10: Averaged RMS error for tone 0 of male speech

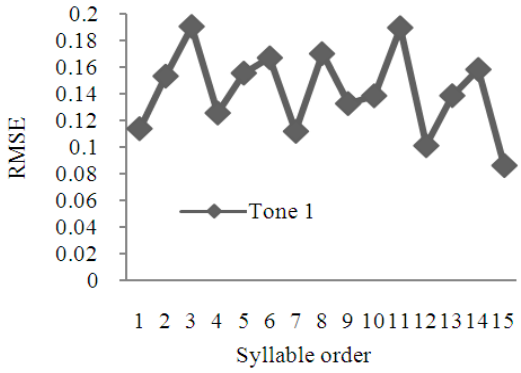

Fig. 11: Averaged RMS error for tone 1 of male speech

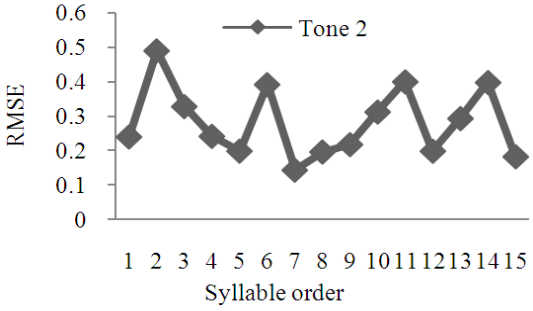

Fig. 12: Averaged RMS error for tone 2 of male speech

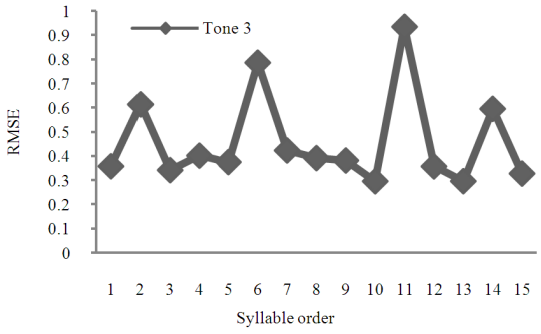

Fig. 13: Averaged RMS error for tone 3 of male speech

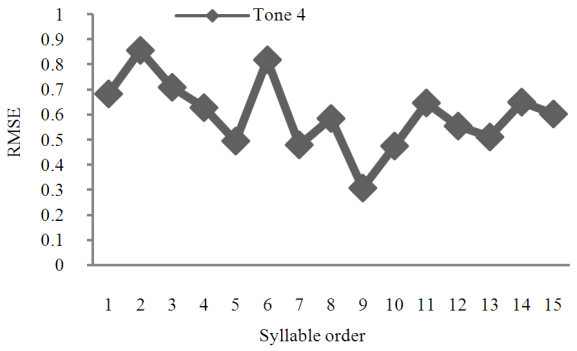

Fig. 14: Averaged RMS error for tone 4 of male speech

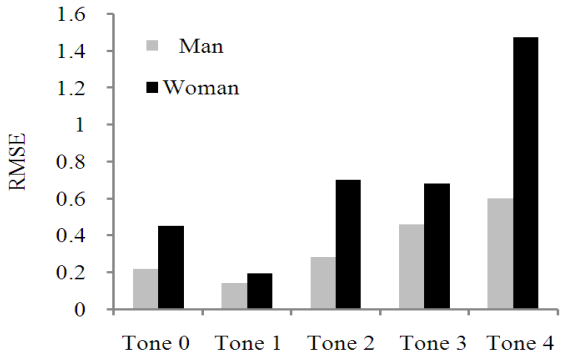

Fig. 15: Comparison of averaged RMS error for five tones of female and male speech

The first five figures (Fig. 5-9) are of female speech. Each figure is of each lexical tone. Meanwhile the next five figures (Fig. 10-14) are of male speech. Finally, Fig. 15 summarizes the averaged value of RMS error of all five tones for both male and female speech.

DISCUSSION

It can be seen from the experimental figures of Figs. 5-14, it has been noticed from both female speech and male speech that the averaged RMS errors of all tones are quite different. Comparing between female speech and male speech in Fig. 15, it can be obviously noticed that the errors of female speech are higher than that of male speech for all tones. Another point, the averaged RMS errors of tone 4 are in the highest level for both female and male speech. Moreover, the averaged RMS errors of tone 1 are in the lowest level for both female and male speech.

CONCLUSION

The structural modeling of fundamental frequency contour for Thai tones is presented in this study. Five lexical Thai tones are statistically studied. The averaged root mean square error of each tone differs from a tone to the others. All in all, the structural modeling technique can be appropriately applied for modeling of Thai tones.

ACKNOWLEDGEMENT

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

REFERENCES

Chomphan, S., 2011a. Analysis of fundamental frequency contour of coded speech based on multi-pulse based code excited linear prediction algorithm. J. Comput. Sci., 7: 865-870. DOI: 10.3844/jcssp.2011.865.870

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. Japan, 57: 445-452.

Fujisaki, H. and S. Ohno, 1998. The use of a generative model of F_0 contours for multilingual speech synthesis. Proceeding of the International Conference on Spoken Language Processing, Oct. 12-16, 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. Proceeding of the International Conference on Spoken Language Processing, (SLP' 90), China, 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 International Conference on Signal Processing, Aug. 26-30, IEEE Xplore Press, Beijing, 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

Ni, J. and K. Hirose, 2006. Quantitative and structural modeling of voice fundamental frequency contours of speech in Mandarin. Speech Comm., 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. Proceeding of the 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. Proceeding of the International Conference on Acoustics, Speech and Signal Processing, May 13-17, IEEE Xplore Press, Orlando, 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. Proceeding of the International Conference on Acoustics, Speech and Signal Processing, Apr. 6-10, IEEE Xplore Press, Hong Kong, pp: 452-455.

Tao, J., J. Yu and W. Zhang, 2006. Internal dependence based f0 model for mandarin tts system. Proceeding of theTC-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. Proceeding of the International Symposium on Tonal Aspects of Languages, (TAL' 06), La Rochelle, France.