Aliasing-free L-F model and its application to an interactive MATLAB tool and test signal generation for speech analysis procedures

Hideki Kawahara[1]

[1]Wakayama University, Wakayama, 640-8510 Japan

kawahara@sys.wakayama-u.ac.jp

Abstract

This demo introduces a closed form representation of the L-F model for excitation source. The representation provides flexible of source parameters in continuous time axis and aliasing-free excitation signal. MATLAB implementation of the model combined with an interactive parameter control and visual and sound feedback is a central component of educational/research tools for speech science. The model also provides flexible and accurate test signals applicable to test speech analysis procedures, such as F0 trackers and spectrum envelope estimator.

Index Terms: glottal source, anti aliasing, speech analysis, speech synthesis, LPC family

1. Introduction

The source-filter model of speech production [1] combined with the one dimensional acoustic tube model of vocal tract [2] still serve as a relevant introduction to speech science. We introduced a closed form representation of an anti-aliased version [3] of the L-F model of voice excitation source [4]. This demo introduces two applications of this aliasing-free L-F model. First one is an interactive MATLAB tool for studying speech production basics, SparkNG [5]. Second application is test signal generation for evaluating speech analysis procedures, such as F0 tracking [6] and spectrum envelope extraction. The following sections briefly describe these applicatoins.

2. SparkNG: interactive MATLAB tools

The speech production simulator of SparkNG has two GUIs. The main GUI is for designing vocal tract shape and transfer function and synthesizes speech sounds using the designed vocal tract and the source signal. The source model GUI is for designing glottal source model based on the L-F model. These GUIs call elementary functions. They consist of conversion functions of LPC family oarameters [7], anti-aliased L-F model signal generator [3], lattice filter based on PARCOR and so on. Figure 1 shows the speech production simulator GUIs.

3. Application: test signal generation

The spurious level other than harmonic component is attenuated more than 120 dB around the fundamental component. These make this model output as an ideal signal for testing tracking ability of F0 extractors. This model is useful to extend a new framework for profiling F0 extractors [8]. A new F0 extraction framework [6] used this model to generate test signals for evaluating its tracking ability. Using a temporally variable lattice filter in the constituent functions also makes quantitative analysis of the time varying group delay effects.

4. Conclusions

This demo introduces application of the aliasing-free L-F model. First one is a set of interactive tools for studying fundamentals of speech production, perception and processing. Second one is test signal generation for speech analysis procedure evaluation. All these are open sourced hoping for them to be useful for beginners as well as experienced tutors and researchers.

5. Acknowledgements

This work was partly supported by Kakenhi (Grant in Aids for Scientific Research) B 15H02726 and 16K12464 of JSPS.

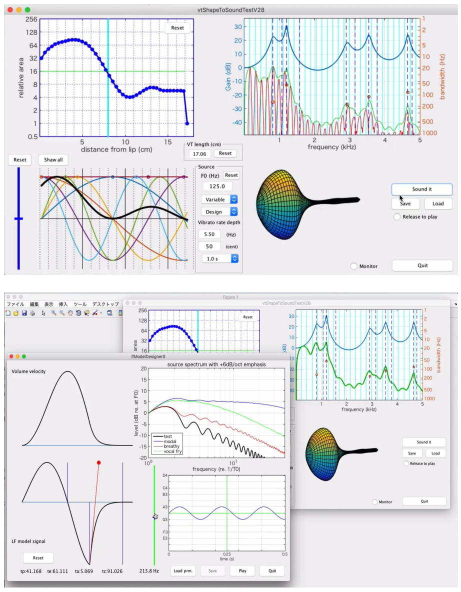

Figure 1: Main GUI of the speech production simulator (upper) and excitation source GUI (lower).

6. References

[1] G. Fant, *Acoustic Theory of Speech Production*. The Hague, The Netherlands: Mouton, 1960.

[2] J. L. Kelly and C. C. Lochbaum, "Speech synthesis," *4th International Congress on Acoustics*, p. G42, 1962.

[3] H. Kawahara, K.-I. Sakakibara, H. Banno, M. Morise, T. Toda, and T. Irino, "Aliasing-free implementation of discrete-time glottal source models and their applications to speech synthesis and F0 extractor evaluation," in *APSIPA 2015*, Hong Kong, 2015, pp. 520–529.

[4] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *Speech Trans. Lab. Q. Rep., Royal Inst. of Tech.*, vol. 4, pp. 1–13, 1985.

[5] H. Kawahara, "Matlab realtime speech tools and voice production tools." [Online]. Available: http://www.wakayama-u.ac.jp/%7ekawahara/MatlabRealtimeSpeechTools/

[6] H. Kawahara, Y. Agiomyrgiannakis, and H. Zen, "Using instantaneous frequency and aperiodicity detection to estimate F0 for high-quality speech synthesis," in *ISCA SSW9*, San Francisco, 2016, [Submitted]. [Online]. Available: http://arxiv.org/abs/1605.07809

[7] S. Sagayama and F. Itakura, "Symmetry between linear predictive coding and composite sinusoidal modeling," *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, vol. 85, no. 6, pp. 42–54, 2002.

[8] M. Morise and H. Kawahara, "TUSK: A framework for overviewing the performance of F0 estimators," in *Interspeech 2016*, San Francisco, 2016, [Accepted].