Predicting single-word F0 contours of Tokyo Japanese using PENTAtrainer2

Albert Lee[1], Yi Xu[1], and Santitham Prom-on[2]

([1]University College London/[2]King Mongkut's University of Technology Thonburi)

This study is a user report on PENTAtrainer2 (Prom-on & Xu, 2012), a semi-automatic software package that extracts the underlying representation of communicative functions of any language. The corpus used consists of 2640 utterances of single words framed in a carrier sentence, controlling for word length, accent condition, and syllable structure. By feeding the voice data and their corresponding annotation into the program, three parameters, namely height (*b*), target slope (*m*), and strength of target approximation (λ), are learned for each condition as defined by the annotator. Using these extracted values, F0 contour is generated and compared with individual real utterances in the corpus.

Accuracy of the predicted F0 contour is assessed in terms of RMSE and Pearson's *r*. The learning process was replicated under various annotation schemes (two tones vs. three tones, syllable vs. mora segmentation), but results were very similar among different schemes. Accuracy was high for the accented stimuli (*r*>0.94) whereas for unaccented words accuracy was not as satisfactory (*r*=0.598, s.d.=0.279). Accuracy was not very different among different word lengths and syllable structures. These results set a good foundation for future F0 synthesis attempts of longer utterances.

REFERENCE

Prom-on, S., & Xu, Y. (2012). PENTATrainer2: A hypothesis-driven prosody modeling tool. In *Proceedings of the 5th IESL Conference on Experimental Linguistics* (pp. 93–100). Athens, Greece.
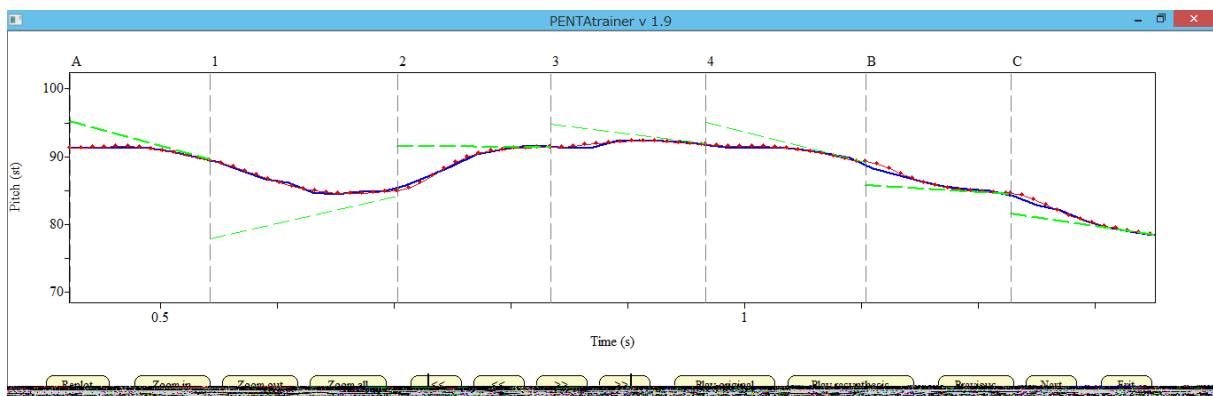
Fig. 1. Original vs. synthesized F0 contour of 'Jiten-**ni nama`nama-mo no**ttemasu 辞典に「生生」も載ってます' (underlined portion displayed)
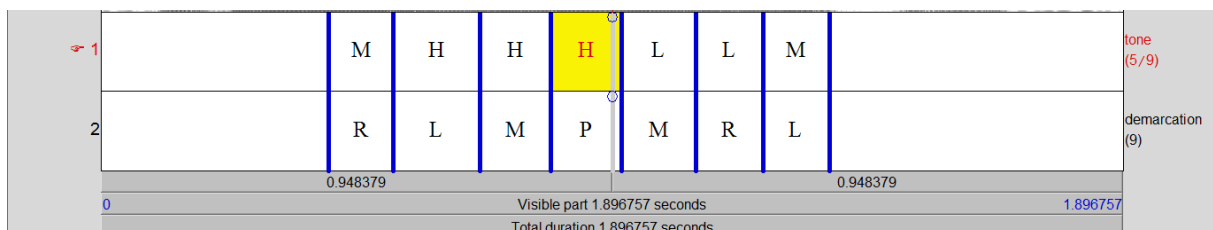

Fig. 2. Example of syllable-by-syllable annotation. Only two layers are used in this study.