The Roles of a Generative Model in the Study of Tonal Features of Speech

Hiroya Fujisaki, Prof. Emeritus, University of Tokyo
fujisaki@alum.mit.edu

Abstract

This paper presents the author’s view on the roles of models in the study of speech, with particular emphasis on generative and quantitative models in studying tone and intonation, and explains the importance of physiology, physics and mathematics in constructing such models. It then describes the author’s own approach in building a model for the process of generation of the contour of the fundamental frequency (henceforth the $F_0$ contour) of speech. Finally, it illustrates the use of the model in various aspects of speech research.

In the study of phonetic and phonological aspects of prosody, the model provides a means to analyze an observed $F_0$ contour of an utterance into its constituent components, and allows one to define the units and structures of prosody, not only of words, phrases, and sentences, but also of larger units of spoken language. This will be demonstrated by a study of prosodic organization of Japanese text reading.

In speech analysis and synthesis, the model provides a means to generate extremely close approximations to observed $F_0$ contours from a limited number of commands that broadly correspond to the lexical and syntactic information of the corresponding text, but not entirely constrained by the text. This point is illustrated by examples from many languages. A conspicuous advantage of the model over other models is that $F_0$ contours generated by the model retain their naturalness even when they are not close approximations to observed $F_0$ contours. This is because the model incorporates the constraints imposed by the physiological and physical properties of the mechanisms of laryngeal control.

In speech recognition, the model can be used to provide information on the syntactic structure and the lexical constituents of an utterance, and thus to improve the recognition performance that is obtained by using only segmental information.

Finally, the model provides a novel means to look at the typological differences and the phonological structures of tone systems of various tone languages. This point will be discussed only briefly here, but will be described in more detail in a separate contributed paper at this symposium.