

Functional annotation for prosodic synthesis

Yi Xu¹ & Santitham Prom-on^{1,2}

¹Department of Speech, Hearing and Phonetic Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, UK.

²Department of Computer Engineering, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

It is widely assumed that prosodic annotation systems are analogous to IPA. It is rarely questioned, however, why IPA does not annotate spectral patterns such as formant, frication or release burst, while prosodic transcription has to annotate pitch patterns such as H, L, up, down, turning point, downstep, etc. Such acoustic transcription makes even less sense when the annotation is used in computational modeling, where the acoustic parameters are automatically learned during model training anyway. As an alternative, we present a new paradigm in which only the temporal intervals of components of hypothetical prosodic functions are labeled, while the discovery of the function-specific parameters is left to the model training process. We tested the paradigm on the *quantitative Target Approximation* (qTA) model, which was implemented in PENTAtainer as a Praat script. The script used a stochastic method to optimize categorical target parameters for each prosodic function across an entire training set. The functional parameters were then used in prosody synthesis and checked against real speech data. Initial results in terms of RMSE, correlation and perceptual impression are highly encouraging. Because of its flexibility, the paradigm is potentially also applicable with other computational models of speech prosody.

Index Terms: prosody annotation, communicative functions, PENTA, PENTAtainer, target approximation