Assessing tonal specifications on a token-by-token basis

Shigeto Kawahara (Keio), Jason Shaw (Yale) & Ishihara Shinichiro (Lund)

BACKGROUND: The current project is driven by two research questions: (1) how can we determine whether a particular speech sample has an F0 target or, alternatively, that F0 is due to interpolation between flanking targets, and (2) how can we analyze individual intonational contours without averaging across multiple speakers, items, or repetitions. The first question is difficult to address, as while F0 contours with unspecified targets generally look like roughly "linear" interpolations, the signal is often noisy and never completely linear. The second question is important, because averaging may miss important inter-speaker or inter-repetition differences. This paper reports a computational toolkit that assigns a probability of tonal specification on a token-by-token basis.

METHOD: The empirical focus of the current study is realization of lexical LHL tonal accent following wh-elements in Tokyo Japanese. Deguchi & Kitagawa (2002) argue that lexical accent after wh-elements in Japanese is "eradicated,", e.g., but other studies cast doubt on this claim (Hirotani 2005; Maekawa 1994). The current study thus reanalyzes a subset of the data obtained by Ishihara (2011) on a token-by-token basis. We compare two sentence structures in (1) and (2).

(1) Control sentences: Word_1 $\text{Word}_{2[\text{-wh}]}$ Word_3 Word_4 Verb
(2) Test sentences: Word_1 $\text{Word}_{2[\text{+wh}]}$ Word_3 Word_4 Verb

The sentences in (1) serve as the control sentences in which the lexical accents of both Word_3 and Word_4 are realized (i.e. full target). We are interested in whether Word_3 and Word_4 in (2) maintain some traces of lexical accent or whether the accent is completely eradicated. There were six types of sentences for both (1) and (2); nine native speakers of Tokyo Japanese repeated those sentences, together with other sentences, two times each. The intonational contours of Word_3 and Word_4 are delimited by their L tones. Following Shaw & Kawahara (2018), these contours were decomposed into the sum of cosines, using DCT. Four DCT coefficients were sufficient to explain > 90% of the variance in F0 trajectories. The variability around the cosine coefficients fit to the data in (2) was used to stochastically simulate linear trajectories between the two L tones. F0 contours simulated from this realistically noisy linear interpolation along with trajectories from the control sentences served as training data for a Bayesian classifier. The classifier assigned a posterior probability of linear interpolation to each token of a test sentence (2).

RESULTS: For Word_3, many speakers produced tokens that have a high posterior probability of linear interpolation. However, some speakers show a large number of tokens that are better classified as belonging to the full target category (i.e. no trace of reduction). We also observe tokens whose posterior probabilities are in the middle range; these tokens are phonetically reduced. For Word_4, most speakers produced tokens assigned a high probability of linear interpolation. Many speakers also produced tokens with full tonal targets and there are many tokens that are phonetically reduced. For both Word_3 and Word_4, we observe both inter- and intra- speaker variability.

DISCUSSION: Our computational toolkit allows us to assess the presence of intonational targets on a token-by-token basis, disentangling phonetic reduction from categorical absence of tonal specification. Most previous studies analyze averaged contours, but analyzing only averaged contours can be misleading. We observe, for example, that one speaker shows reduction for all tokens, whereas another speaker shows a bimodal distribution of full targets and eradication. If we were to be only looking at averages, we would have erroneously concluded that both speakers show reduction. This highlights the importance of analyzing each token separately. Second, several previous studies used a linear regression analysis to assess a linear interpolation analysis (Maekawa 1994; Pierrehumbert & Beckman 1988). One advantage of our approach is that it does not assume linearity in the signal; this is because DCT, the first step of our analysis can represent any trajectory with an arbitrary level of precision. Overall, we conclude that token-by-token analysis offers great promise for the study of intonational variation.

Shaw, J. A., & Kawahara, S. (2018). Assessing surface phonological specification through simulation and classification of phonetic trajectories. *Phonology, 35*(3), 481-522. doi:10.1017/S0952675718000131