Contextualized embeddings predict the pitch contours of Mandarin monosyllabic words

Xiaoyun Jin, Rolf-Harald Baayen

We present a corpus-based investigation of how the pitch (F0) contours of mono-syllabic words are realized in spontaneous conversational Taiwan Mandarin, focusing on possible effects of words' meanings on phonetic realization. Our hypothesis is meaning co-determines pitch contours, which is motivated by prior research on English homophones, the duration of English /s/, and pitch contours of disyllabic Mandarin words.

In the present study, we analysed the pitch contours of monosyllabic Mandarin words. We used the Generalized additive model (GAM) to analyse the pitch contours of 6118 tokens representing 94 word types, which jointly represent 210 word senses. Even after controlling for the variables such as word duration, gender, speaker identity, tonal context, vowel height and utterance position, both word identity and word sense emerged as significant predictors of tonal realization.

For heterographic homophones in conversational Mandarin, differences in phonetic realization are present for pitch. For 你 and 妳 (the male and female second person personal pronouns), 妳 tends to have higher pitch height than 你. Likewise, 拿 (roughly similar to English 'take') has tonal contours that depend on word sense. For example, when 拿 is used with the sense of using the hand to take or hold something (e.g., 拿菜單, 'take the menu') its pitch contour differs from that realized for the more abstract sense 'take somebody as an example' (e.g.,拿他來舉例, 'take him as an example').

More generally, we are finding that words that according to standard dictionaries supposedly have the same tone, are actually realized in conversational speech with their own tonal signatures. The tonal signatures of individual word tokens turn out to be predictable far above chance level from their corresponding contextualized embeddings, obtained using a Mandarin GPT-2 model. Given that many of the monosyllabic words in our study are highly-polysemous function words, this is a non-trivial result. Our results provide further evidence for the existence of isomorphies between word meaning in context and phonetic realization.