

Perception of intonation in Cantonese: Native listeners versus exemplar-based model

Una Y. Chow, Stephen J. Winters
University of Calgary

This study investigated the perception of Cantonese intonation by an exemplar-theoretic model, without using F0 normalization. Exemplar Theory (Johnson, 1997) claims that listeners store experienced utterances (i.e., exemplars) in rich phonetic detail, so that when a new utterance is heard, it is classified by comparing its auditory similarity to all exemplars stored in memory, without using speaker normalization. Chow and Winters (2015) applied Johnson's model to intonation perception and demonstrated that the model could correctly classify 95% of statements and questions in Cantonese, based on F0 cues alone. Cantonese provides a challenging test case for this model, because its high boundary tone on echo questions (Wong *et al.*, 2005) potentially creates perceptual confusion with its rising tones, 25 and 23, on the final syllable (Ma *et al.*, 2006). In this study, we compared the performance of human listeners in a perception study to the performance of this computer model, in order to determine whether an exemplar-based model could account for human perception of intonation in statements and questions in Cantonese.

Chow and Winters (2015) recorded five male and five female native speakers from Hong Kong producing twenty different pairs of statements and echo questions in Cantonese. First, we randomly selected the utterances of two male and two female speakers for this study. Then we presented these 160 recorded sentences to twenty native listeners of Cantonese in an identification task in three different forms: 1) the complete sentence (e.g., *Wong⁵⁵ Ji²² gaau³³ lik²².si²⁵* 'Wong Ji teaches history'), 2) the non-final portion of the utterance (e.g., *Wong⁵⁵ Ji²² gaau³³ lik²²* 'Wong Ji teaches his...'), and 3) the final syllable of the utterance (e.g., *si²⁵* '...tory'). The listeners were trained on 50% of the sentences and then tested on the remaining 50%. Finally, we trained and tested the model in the same conditions as in the training and testing of the human listeners. The model categorized statements and questions based on the auditory similarity between a new token and each of the previously presented tokens, calculated by applying an exponential function to the Euclidean distance between F0s of eleven equidistant time points of the periodicity of the compared tokens.

ANOVAs on perceptual sensitivity (d') and response bias (β) revealed that both listeners and the model were significantly less sensitive to non-final stimuli than to complete sentences and final syllables ($p < 0.05$). However, listeners were able to identify statements and questions from non-final stimuli significantly better than the model, but showed significantly more bias towards statements in this condition ($p < 0.05$). These results suggest that native listeners rely primarily on F0 cues in the final syllable to identify statements and questions. In the absence of these F0 cues, listeners tended to identify non-final stimuli as statements. Both listeners and the model performed well above chance on final syllables. There was no significant difference between their performances in this condition. The results of this study show that an exemplar-based model, without F0 normalization for speakers, is a promising model for human intonation perception of statements and questions in a language that relies primarily on F0 for both lexical tones and sentence intonation, such as Cantonese.

References

Chow, U. Y., & Winters, S. J. (2015). Exemplar-based classification of statements and questions in Cantonese. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-165). San Diego: Academic Press.

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *Journal of Acoustical Society of America* 120(6), 3978-3987.

Wong, W. Y. P., Chan, M. K. M., & Beckman, M. E. (2005). An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 271-300). New York: Oxford University Press.