An exemplar-based learning model of English sentence intonation

Una Y. Chow & Stephen J. Winters
School of Languages, Linguistics, Literatures and Cultures, University of Calgary

This study investigated whether an exemplar-based model could learn to categorize English statements and echo questions based on intonation, without normalizing fundamental frequency (F0) to account for within-speaker and between-speaker variability. Exemplar theory (Johnson, 1997) claims that listeners store exemplars of speech that they have experienced in rich phonetic detail, which allows them to categorize new words in memory without speaker normalization. In applying exemplar theory to intonation perception, we have developed a computational model that categorizes English statements and echo questions by comparing how similar a sentence is to previously encountered sentences, based solely on intonation. This model adapts and applies Nosofsky's (1988) and Johnson's (1997) algorithm for calculating similarity between sentences.

In English, the pitch of the voice tends to fall at the end of a statement but tends to rise at the end of an echo question (Wells, 2006). The point where this particular pitch begins to fall or rise in the sentence is the nuclear tone. Therefore, our model calculates similarity using the following three auditory properties: 1) the timing of the nuclear tone in the sentence, 2) the speed of the F0 change at the nuclear tone, and 3) the direction of F0 change at the nuclear tone.

We designed the model with a computational interface that has six interactive user functions, which allow users to run each function separately and modify the default parameters, if necessary. Example parameters of the model include F0 range, percentages of trained and test data, and learning rate. The *preanalysis* function removes any silence or noise before and after an utterance. The *analysis* function first applies interpolation to the F0 contour of an utterance to create a continuous contour. It then locates the nuclear tone in the utterance (as shown in Figure 1). The *extraction* function extracts the auditory property measurements from the utterances. In calculating similarities, the model assigns different weights to the auditory properties. The *training* function (Figure 2) trains the model to derive a set of generalized weights for the auditory properties that yields an optimal accuracy rate in categorizing new sentences. The *testing* function tests how accurately the model can categorize statements and echo questions from a set of sentences which differs from the training set. It uses the weighted sum of the auditory properties to estimate to which category a new sentence belongs. Finally, the *cross-validation* (Refaeilzadeh, Tang, & Liu, 2009) function evaluates how well the model generalizes to new tokens, with successive runs that will eventually test each token once and only once.

We trained and tested the model on 64 pairs of English statements and echo questions produced by two adult, native speakers (one male and one female). For example, "Ann is a teacher. Ann is a teacher?" The model correctly categorized up to 80-100% of the sentences. The result of the model's performance demonstrated that it is feasible to develop an exemplar-based computational model that can learn to categorize statements and echo questions without normalizing F0 to account for speaker variability.

References

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.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 700-708.

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In L. Liu & M. T. Zsu (Eds.), *Encyclopedia of database systems* (pp. 532-538). Springer Publishing Company Incorporated.

Wells, J. C. (2006). *English intonation: An introduction*. Cambridge: Cambridge University Press.

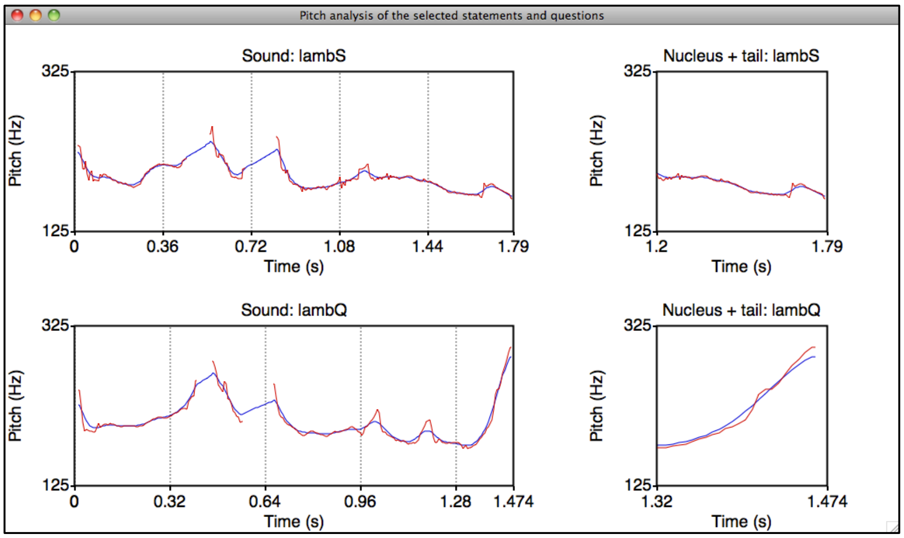
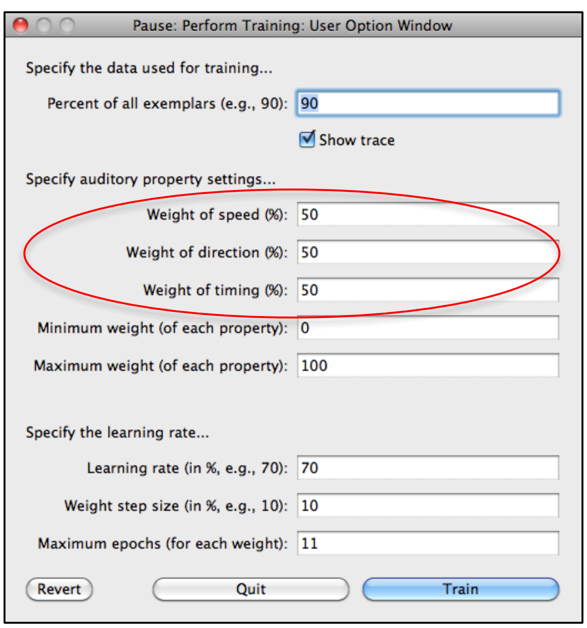


Figure 1. Output of the *analysis* function on "Mary has a little lamb." (top) and "Mary has a little lamb?" (bottom). Complete F0 contours (left); tail ends starting from the nuclei (right). Original F0 contours (red); interpolated contours (blue).



Initial, user-specified weights of the *speed*, *direction*, and *timing* properties, which will be adjusted by the model during the learning process to obtain a generalized set of weights.

Figure 2. The *training* function.