Perceiving Tone Categories of Shanghainese
—an Experimental Study and Computational Simulation

Yuanliang Meng, University of Kansas, ymeng@ku.edu

Tones in Shanghai Chinese are not fully independent phonological categories. Instead, they co-vary with the features of syllable onset, such as voicing and duration, and these features are context-dependent. This research investigates two tones, namely *yinqu* (T2) and *yangqu* (T3). The project has three parts: (1) acoustic experiment, (2) perceptual experiment and (3) computational simulation. We claim that the identification of these 'tones' involves the perception of a set of fine-grained features, and learning the phonological contrast requires mapping the features to various contexts. As a simplified model, three dimensions of features need to be considered for a single phonological contrast. This can be captured easily by an artificial neural network. In the network, the weight of each feature is learned from language input and is relatively stable. Altering the features may or may not change the perception significantly, which is predicted by the model.

Acoustic experiment: We recruited 11 native speakers and recorded their pronunciations of words containing T2 or T3. Pitch height, onset duration, and degree of voicing of the target onsets were measured. Paired t-tests show significant differences in relevant conditions and the results agree with previous phonetic studies. In monosyllabic words, T2 is realized as a high rising tone, while T3 is a low dipping tone. The syllable onset is typically voiceless in both cases, with a positive VOT. In word-medial position, however, T2 syllables have relatively long, voiceless onsets while T3 syllables have short, voiced onsets. Depending on word stress pattern, the pitch contours of the tones could undergo neutralization as well, but the contrast in onset retains.

Perceptual experiment: The stimuli included both real words and artificially manipulated words. Specifically we swapped the tones or onsets between paired-words, and adjusted the duration of onsets. 50 native speakers were recruited. After each stimulus was presented, they picked what they heard from two choices. The subjects reached high accuracy for control stimuli (real words). Manipulating acoustic features in certain conditions led to a variety of perceptual results, implying the weights of features.

Computational simulation: A feedforward artificial neural network is used for simulation. It has three input nodes, one output node and a hidden layer. Each input node accepts a feature and the output node delivers a value between 0~1. We trained the network with discrete values and continuous values, respectively. Testing data represent both real productions and the artificially manipulated productions; and the results are compared with those from perceptual experiment. Essentially, our preliminary results have revealed that human subjects and the artificial neural network behave similarly on innovative data.