Auditory and Visual Cues in the Production and Perception of Mandarin Tones

Yueqiao Han[1], Thiago Castro Ferreira[2], Martijn Goudbeek[1], Maria Mos[1], and Marc Swerts[1]

[1]Tilburg University
[2]Federal University of Minas Gerais, UFMG

Which acoustic and visual cues can be used to classify Mandarin tones? Furthermore, are these cues the same for producers (i.e., the ground truth) and (tone-naïve) perceivers? To address these questions we video-taped four Mandarin speakers while they produced ten syllables with four Mandarin tones, i.e. forty words in two styles (natural and teaching), totaling 160 stimuli. The audiovisual stimuli were subsequently presented to 43 tone-naïve participants in a tone identification task. Basic acoustic and visual features were extracted. We used decision trees and machine learning to identify the most important acoustic and visual features for classifying the tones. These features were identified separately for the tones as produced by the four speakers and for the tones as identified by the 43 perceivers. Based on previous work, we expected that acoustic features would be more relevant than visual features, but that non-native perceivers might still benefit from the visual signal. The results showed that acoustic features were ranked higher than the visual features for tone classification, both for the classification of the intended and the perceived tone. However, tone perceivers did revert to the use of visual information in certain cases. So, while visual information does not seem to play a significant role in native speakers' tone production, tone-naïve perceivers do sometimes consider visual information in their tone identification.