Different facial cues for different speech styles in Mandarin tone articulation.

Saurabh Garg¹, Lisa Tang², Ghassan Hamarneh³, Allard Jongman⁴, Joan Sereno⁴, Yue Wang¹

1. Language and Brain Lab, Department of Linguistics, Simon Fraser University, Burnaby, BC, Canada.
2. Digital Emergency Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada.
3. Medical Image Analysis Lab, School of Computer Science, Simon Fraser University, Burnaby, BC, Canada.
4. Phonetics & Psycholinguistics Laboratory, Department of Linguistics, University of Kansas, Lawrence, KS, United States.

ABSTRACT BODY:

Abstract (200 words): Research has shown that facial articulatory cues aid speech perception. However, how such cues are employed in different speech styles remains unclear. This study examined facial articulatory features of Mandarin tones in clear versus conversational speech styles produced by 20 native Mandarin speakers. Using computer-vision and image-processing techniques, keypoints representing each speaker’s head, eyebrow and lips were identified on video, and their movement trajectories during tone productions were tracked. Thirty-three features based on distance, time and kinematics (e.g., velocity) were subsequently computed to characterize the movements. Random forest and t-test analyses were then conducted to identify the significant features between the two speech styles. Results reveal that, across tones, clear relative to conversational style involves greater movement distance and velocity, reaching articulatory targets faster. Individual tone analyses further indicate that the faster target approximation in clear speech occurs for contour tones (2-4) but not the flat tone 1. The increase in distance and velocity in clear speech is reflected on more features for the most dynamic tone 3 than for the other tones. These results suggest that clear-speech modifications for tones can be exhibited through facial movements, involving hyper-articulation across tones and tone-specific adjustments aligned with individual tone trajectories.