Perception of Intonational Cues in Speech-act Intentions in Cantonese by Native Speakers

Authors: Meixuan Li, The Hong Kong Polytechnic University
Si Chen, The Hong Kong Polytechnic University
Abstract ID: 348
Event: Phonetics 2026 Hong Kong
Subject: 3. Speech perception
Presenter Name: Meixuan Li | Si Chen

Objectives

This study probed whether native listeners could perceive speech-act meaning from prosody in the absence of lexical or contextual support. Specifically, we asked (i) how accurately adult Cantonese speakers discriminate six illocution types—Statement, Doubt, Suggestion, Command, Celebration and Complaint; (ii) whether accuracy is modulated by utterance length, comparing five-syllable sentences with isolated disyllabic words; and (iii) what systematic confusions reveal about the distinctiveness of the corresponding intonational profiles.

Methodology

Twenty neurotypically developed native Cantonese speakers (10 male, 10 female; aged 21–27; normal hearing) completed a six-alternative forced-choice task. Stimuli were produced by two 21-year-old voice actors (1 male, 1 female): 9 disyllabic words and 24 neutral five-syllable sentences, each recorded with six speech-act intentions and repeated three times, yielding 324 word tokens and 864 sentence tokens. Stimuli were presented in random order over headphones in a sound-proof booth, and participants identified the intended speech act by key press. Accuracy was analyzed using a binomial generalized linear mixed-effects model with fixed effects of Illocution Type and Domain (Word vs. Sentence), and random intercepts for Participant and Item. Tukey-adjusted pairwise contrasts were computed with emmeans.

Results

Overall accuracy reached 62.2 % (95 % CI = 61.3–63.1), far surpassing the 16.7 % chance baseline ($\chi^2 > 1000$, $p < .001$; $\varkappa = 0.55$). Per-category sensitivities showed highest recoverability for Celebration (0.87) and Complaint (0.90), intermediate performance for Doubt (0.69) and Suggestion (0.57), and lowest for Statement (0.51) and Command (0.44), the latter two being mutually confusable. Mixed-model estimates, referenced to sentence-level Statements, confirmed that all other forces except Doubt were significantly harder to identify (Δlog-odds = –1.50 … –3.00, $|z| \geq 8.6$, $p < .001$). A main effect of Domain indicated that Statements were easier to recognise when realised as isolated words ($\beta = 0.89$, $z = 2.41$, $p = .016$). Crucially, a Domain × Illocution interaction showed that truncation to the word level significantly diminished recognition of Doubt ($\beta = –1.75$), Command ($\beta = –1.90$) and Celebration ($\beta = –0.85$) but left Suggestion and Complaint unaffected ($|z| \geq 2.39$, $p \leq .017$). Tukey contrasts revealed that within sentences, Statements outperformed all forces except Doubt (OR = 1.64, $p = .064$), whereas within words they exceeded every other category by large margins (ORs = 9.48–61.62, $p < .001$).

Discussion

The present findings demonstrate that even under the informationally impoverished conditions, Cantonese provides listeners with a prosodic inventory rich enough to convey at least a rough map of speech-act meaning. This challenges the assumption that a dense lexical-tone system necessarily undermines higher-level intonation. Instead, our 62 % accuracy rate, coupled with a substantial Cohen's $\varkappa$ (.55), shows that lexical tone and sentence intonation can exist in the same acoustic channel.

Within autosegmental–metrical (AM) models, this suggests that Cantonese possesses dedicated edge tones or phrase-level configurations that are functionally independent of the lexical tone tier, warranting a refinement of current tone–intonation interface proposals.