Acoustic features of Cantonese speech acts: prosodic evidence from words and sentences

Meixuan Li, Bingxin Liu, Si Chen
The Hong Kong Polytechnic University, Hong Kong

https://doi.org/10.36505/TheLinguisticProceedings/2025/16/01/016/000676

Abstract

Prosodic cues provide a suprasegmental access to speech-act perception, yet they remain under-explored in tonal languages. The present study analysed the production of six speech acts (Statement, Doubt, Suggestion, Command, Celebration and Complaint) by twenty Cantonese adults. Mean F_0, F_0 range, intensity and duration jointly doubled chance-level classification, despite competition from lexical tone. Doubt and Suggestion exhibited the highest mean F_0 and widest span; Statement and Complaint occupied the lowest pitch region; Command showed the narrowest span and greatest loudness; Complaint was longest in duration. The findings confirm that global intonation, intensity and timing reliably encode pragmatic force in Cantonese.
Keywords: Cantonese, speech act, prosody, production

Introduction

Language can do things. When words are uttered, the intentions of the speaker are implied and inferred by the listener in real time. According to Speech-Act theory (Austin 1962; Searle 1969), this inference relies on illocutionary-force-indicating devices distributed across the lexical, syntactic and prosodic tiers. Among these, prosody is uniquely efficient with its independency. Previous studies have shown that global F_0 height, contour shape and boundary direction map systematically onto speech-act categories (Hellbernd & Sammler 2016). Cantonese complicates the picture because the same F_0 channel is already committed to lexical tone. Production studies on questions, statements and imperatives (Ma et al. 2008) suggest that register shifts and boundary tones can coexist with lexical tone. Yet the empirical base remains fragmentary, limited to isolated contrasts and adult speakers. The present experiment therefore asks: How are six everyday speech acts encoded prosodically in Cantonese, and do global acoustic features differentiate them despite lexical tone?

Hypothesis

Speech acts differ systematically in (i) Mean F_0 & pitch range, (ii) Intensity, (iii) Duration. These cues allow machine classification above chance.

Proceedings Phonetics 2025 Hong Kong: 16th International Conference on Linguistic Research and Applications

Methods

Participants and materials

Twenty Cantonese adults (10 F/10 M; 21–27 yrs; with normal hearing and speech) were recorded. Materials comprised nine disyllabic words covering tone combinations and six carrier sentences "我哋 + Verb-Object". Six acts × two repetitions yielded 180 tokens per speaker. Contextual scenarios were scripted for natural production.

Procedures

Each trial presented a scenario prompt plus an AI partner's turn. The speaker responded with the target utterance in context. Recordings were made in a sound-proof booth with a high-quality microphone.

Acoustic analysis

Manual segmentation was performed by trained phoneticians. F_0 was normalised to semitones per speaker. Extracted measures were Mean F_0, F_0 range, Mean RMS intensity and Duration.

Data analysis

Descriptive statistics (M ± SE) were plotted for each parameter × speech act × utterance type. A jack-knife linear discriminant analysis (LDA) with leave-one-speaker-out cross-validation assessed how well the four-cue prosodic vector predicted speech-act category. Significance against the 16.7 % chance level was evaluated with χ^2 tests.

Results

Descriptive patterns

Descriptive statistics showed intention-specific prosodic patterns in both word and sentence conditions (Table 1&2). Doubt and Suggestion were realized with the highest mean F0 and the widest pitch spans, whereas Statement and Complaint occupied the lowest pitch region and Command exhibited the narrowest span. Command and Celebration were the loudest (≈ 65 dB), Complaint the longest (≈ 0.8 s in words, 1.3 s in sentences), and Command the shortest.

Classification

A jack-knife linear discriminant analysis using only these four cues (mean F0, F0 range, intensity and duration) classified tokens at 35.3 % accuracy for words and 31.8 % for sentences, roughly double the 16.7 % chance level ($\chi^2 > 184$, $p < 10^{-25}$). Commands and Statements were recognized best (≈ 46–54 % correct).

Confusion patterns

Command and Complaint were best recognized. Celebration and Doubt were often confused with others. There was overlap between Doubt–Suggestion–Statement.

Table 1. Acoustic Summary (words).

Speech Act	Duration (ms)	F0 Range (st)	Mean Intensity (dB)	Mean F0 (st)
Statement	646.68	84.21	59.87	172.94
Doubt	694.28	159.16	61.02	211.25
Suggestion	639.10	154.15	59.81	214.99
Command	577.41	105.12	66.22	199.30
Celebration	699.18	120.99	65.10	207.92
Complaint	830.05	99.96	62.04	171.45

Table 2. Acoustic Summary (sentences).

Speech Act	Duration (ms)	F0 Range (st)	Mean Intensity (dB)	Mean F0 (st)
Statement	1070.80	1070.80	1070.80	1070.80
Doubt	81.55	81.55	81.55	81.55
Suggestion	58.79	58.79	58.79	58.79
Command	171.02	171.02	171.02	171.02
Celebration	1164.36	1164.36	1164.36	1164.36
Complaint	159.62	159.62	159.62	159.62

Discussion

Overall, the results show that Cantonese speakers reliably produce pitch height, pitch span, loudness, and timing patterns to convey different speech acts. These global prosodic patterns are strong enough to yield machine classification at twice-chance accuracy, yet not distinctive enough to prevent specific confusions. This is the first attempt to provide a systematic acoustic map of Cantonese speech-act prosody, showing prosodic space available despite the tone system. These findings confirm that global F0, intensity, and durational cues jointly encode basic pragmatic intentions, though not yet at ceiling discriminability.

Limitations

Four global parameters only. Accuracy is modest; dynamic or spectral features may enhance classification.

Conclusions

Six speech acts are reliably differentiated by global prosody. Distinctions are not perfect but systematic. Cantonese prosody extends beyond lexical tone and provides acoustic benchmarks for further study.

Acknowledgements

This study forms part of my PhD project at The Hong Kong Polytechnic University and was supported by the PolyU Research Postgraduate Scholarship (PRPgS). We thank all participants and research assistants for their contribution.

References

Austin, J. 1962. How to Do Things with Words. Oxford University Press.

Green, M. 2000. Illocutionary Force and Semantic Content. Linguistics and Philosophy 23, 435–473.

Hellbernd, N., Sammler, D. 2016. Prosody conveys speaker's intentions: Acoustic cues for speech act perception. Journal of Memory and Language 88, 70–86.

Ma, J. K.-Y., Ciocca, V., Whitehill, T. L. 2008. Acoustic cues for the perception of intonation in Cantonese. Interspeech 2008, 520–523.

Searle, J.R. 1968. Austin on Locutionary and Illocutionary Acts. The Philosophical Review 77(4), 405–424.

Searle, J.R. 1969. Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press.

Searle, J., Vanderveken, D. 1985. Foundations of Illocutionary Logic. Cambridge University Press.

Wu, W.L. 2009. Sentence-final particles in Hong Kong Cantonese: Are they tonal or intonational? Interspeech 2009, 2291–2294.