

Pitch Elevation on Wh-Words in Cantonese Spontaneous Speech

Ziqi Zhang¹, Anqi Xu², Yu-Yin Hsu¹

¹ The Hong Kong Polytechnic University, ² Harbin Institute of Technology, Shenzhen

zhangziqi.bram@gmail.com, xuanqi@hit.edu.cn, yu-yin.hsu@polyu.edu.hk

The interaction between lexical tone and sentence-level intonation remains a central debate in Chinese prosody. In dense tone languages like Cantonese, speakers face a physiological conflict: they must signal interrogative force, typically *via* significant pitch elevation without compromising the F_0 contours essential for lexical distinction [1]. While previous studies predominantly attribute Cantonese interrogative intonation to Sentence-Final Particles (SFPs) or boundary tones [2], the prosodic realization of *wh*-words *in situ* remains under-explored in spontaneous discourse. This study investigates four primary Cantonese *wh*-words: *mat je* (乜嘢, ‘what’), *bin go* (邊個, ‘who’), *dim gai* (點解, ‘why’), and *me* (咩, ‘what’). We propose a *Lexical Hierarchy of Interrogative Prosody*, positing that pitch elevation for focus marking is not a universal grammatical rule applied uniformly to all question words. Instead, we argue that Cantonese employs an item-specific strategy: some *wh*-words carry a heavy prosodic load with significant pitch raising, while others rely exclusively on morphosyntactic markers, exhibiting no acoustic differentiation from declaratives.

We analyzed spontaneous Cantonese speech data from the Common Voice Corpus 21.0 [3], and built an automated Python 3.11 pipeline employed the Montreal Forced Aligner (MFA) [4] for phone-level alignment and pYAAPT [5] for robust pitch tracking. We extracted the mean F_0 of each target *wh*-word from both declarative and interrogative sentences, normalizing values to Z-scores to minimize inter-speaker variability. To rigorously quantify the prosodic cost of interrogation, we fitted Linear Mixed-Effects Models (LMM) with Sentence Type (Declarative or Interrogative) and Syntactic Position (Initial, Medial and Final Position) as fixed effects. By treating declarative realization as the baseline and statistically controlling for positional effects (Declination [6]), this design isolates the elevation of the interrogative pitch, i.e., the residual rise attributable solely to the interrogative function, independent of global intonational trends and physiological factors.

Analysis of 15,664 valid sentences (Table 1) reveals that interrogative marking is strictly lexically conditioned (Table 2). The item *mat je* (乜嘢, ‘what’) functions as the primary prosodic anchor, exhibiting significant elevation ($\beta = +0.203$ SD, $p < .001$). The adverb *dim gai* (點解, ‘why’), despite frequently appearing in the naturally high-pitched sentence-initial position, carries a significant compensatory lift even after accounting for boundary tone ($\beta = +0.128$ SD, $p < .001$). The particle *me* (咩, ‘what’) likewise shows significant elevation ($\beta = +0.122$ SD, $p < .001$), consistent with boundary tone effects. In stark contrast, *bin go* (邊個, ‘who’) does not exhibit significant pitch elevation ($\beta = +0.026$ SD, $p = .466$), maintaining tonal stability and relying entirely on morphosyntactic marking [7].

This study provides empirical evidence for a non-uniform prosodic grammar in Cantonese. The findings challenge the view [8] of interrogative intonation as a global overlay applied indiscriminately across the sentence. Instead, Cantonese speakers utilize localized, item-specific strategies: most of the *wh*-words like *mat je* (乜嘢, ‘what’) perform a higher fundamental frequency, whereas words like *bin go* (邊個, ‘who’) maintain absolute tonal stability, offloading the functional load of interrogation to syntax.

Table 1: Detailed Positional Distribution of Target Wh-Words by Sentence Type

Target Word	Type	Initial Pos.	Medial Pos.	Final Pos.	Total N
<i>me</i> (咩, ‘what’)	Decl.	253 (5.6%)	2,879 (64.0%)	1,364 (30.3%)	4,496 (58.0%)
	Ques.	134 (4.1%)	1,802 (55.3%)	1,321 (40.6%)	3,257 (42.0%)
	<i>Total</i>	<i>387 (5.0%)</i>	<i>4,681 (60.4%)</i>	<i>2,685 (34.6%)</i>	<i>7,753 (100%)</i>
<i>mat je</i> (乜嘢, ‘what’)	Decl.	151 (8.7%)	1,290 (74.7%)	287 (16.6%)	1,728 (59.2%)
	Ques.	166 (13.9%)	743 (62.4%)	281 (23.6%)	1,190 (40.8%)
	<i>Total</i>	<i>317 (10.9%)</i>	<i>2,033 (69.7%)</i>	<i>568 (19.5%)</i>	<i>2,918 (100%)</i>
<i>dim gai</i> (點解, ‘why’)	Decl.	781 (51.8%)	665 (44.1%)	62 (4.1%)	1,508 (55.6%)
	Ques.	744 (61.8%)	460 (38.2%)	0 (0.0%)	1,204 (44.4%)
	<i>Total</i>	<i>1,525 (56.2%)</i>	<i>1,125 (41.5%)</i>	<i>62 (2.3%)</i>	<i>2,712 (100%)</i>
<i>bin go</i> (邊個, ‘who’)	Decl.	335 (27.0%)	737 (59.3%)	170 (13.7%)	1,242 (61.1%)
	Ques.	218 (27.6%)	464 (58.7%)	109 (13.8%)	791 (38.9%)
	<i>Total</i>	<i>553 (27.2%)</i>	<i>1,201 (59.1%)</i>	<i>279 (13.7%)</i>	<i>2,033 (100%)</i>

Table 2: Linear Mixed-Effects Model Results: Pitch Elevation (Z-Score) in Interrogatives

Target Word	Z-Score Coef.	p-Value
<i>mat je</i> (乜嘢, ‘what’)	+0.203	< .001
<i>dim gai</i> (點解, ‘why’)	+0.128	< .001
<i>me</i> (咩, ‘what’)	+0.122	< .001
<i>bin go</i> (邊個, ‘who’)	+0.026	.466

References

- [1] Joan K-Y Ma, Valter Ciocca, and Tara L Whitehill. “Effect of intonation on Cantonese lexical tones”. In: *The Journal of the Acoustical Society of America* 120.6 (2006), pp. 3978–3987.
- [2] Wing Li Wu. “Cantonese prosody: Sentence-final particles and prosodic focus”. PhD thesis. University College London, 2013.
- [3] Rosana Ardila et al. “Common Voice: A Massively-Multilingual Speech Corpus”. In: *Proceedings of the 12th Language Resources and Evaluation Conference (LREC)*. 2020, pp. 4218–4222.
- [4] Michael McAuliffe et al. “Montreal forced aligner: Trainable text-speech alignment using Kaldi”. In: *Inter-speech*. Vol. 2017. 2017, pp. 498–502.
- [5] Stephen A Zahorian and Hongbing Hu. “A spectral/temporal method for robust fundamental frequency tracking”. In: *The Journal of the Acoustical Society of America* 123.6 (2008), pp. 4559–4571.
- [6] Yi Xu. “Effects of tone and focus on the formation and alignment of F_0 contours”. In: *Journal of Phonetics* 27.1 (1999), pp. 55–105.
- [7] Yu-Yin Hsu and Anqi Xu. “Wh-indeterminates and Prosody in Hong Kong Cantonese”. In: *Proc. 10th International Conference on Speech Prosody*. 2020, pp. 376–380.
- [8] X. N. S Shen. “The prosody of Mandarin Chinese”. In: *University of California Publications in Linguistics* 118 (1990).