THE PROSODIC PROPERTIES OF THE CANTONESE SENTENCE-FINAL PARTICLES *aa1* AND *aa3* IN RHETORICAL *WH*-QUESTIONS

Roger Yu-Hsiang Lo[1], Angelika Kiss[2], and Maxime Tulling[3]

[1]University of British Columbia, [2]University of Toronto, [3]New York University
[1]roger.lo@ubc.ca, [2]angelika.kiss@mail.utoronto.ca, [3]mat646@nyu.edu

ABSTRACT

In this production experiment we investigated the suprasegmental properties of the Cantonese sentence-final particles (SFPs) *aa1* and *aa3* in different types of questions. The context of each trial was manipulated to elicit a set of string-identical *wh*-questions with the following four readings: information-seeking questions (ISQs), positive rhetorical questions (RQ+s) *'Who likes coffee? - John'*, negative rhetorical questions (RQ-s) *'Who likes coffee? - Nobody'*, and rhetorical questions as retort (Retort) *'Who ordered an extra large coffee? - Who likes coffee?'*. Overall, the results show that the prosodic difference in SFPs helps to distinguish between these question types. We found that (i) SFPs in RQ+ and Retort contexts are characterized by rising tonal contours, while those in ISQ and RQ- contexts have more level contours, and that (ii) the duration of SFPs in RQ- contexts are significantly longer than in any other context. We also found that *aa1* and *aa3* pattern similarly.

Keywords: Cantonese, sentence-final particles, rhetorical questions, prosody

1. INTRODUCTION

1.1. Subtypes of rhetorical questions

Most typically, *wh*-interrogatives function as information-seeking questions (ISQs), with the speaker requesting information from the addressee [10]. Rhetorical questions (RQs) differ from ISQs in the sense that the speaker, rather than requesting information, instead strongly suggests a certain answer to the question asked. For instance, the question in (1) can be asked in an unbiased way when the speaker does not prefer any member of the set denoted by the *wh*-word *who*, or in a biased way when the speaker suggests a certain answer to *who*.

(1) Who would drink coffee?

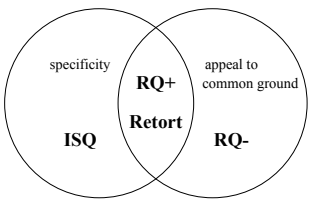
In negative rhetorical questions (RQ-s) the answer suggested by the speaker is the empty set (i.e. 'no one'), providing (1) with a reading equivalent to the speaker saying 'no one would drink coffee' [8].

However, RQs can also suggest a non-empty set as their answers if there is a particular individual salient enough in the discourse context to serve as an answer to the question [19, 3]. These we call positive rhetorical questions (RQ+s). A subtype of RQ+s is positive RQs-as-retorts (Retorts), which are RQ+s used to answer an ISQ. Stylistically, this has the effect of the speaker marking the ISQ as having the same obvious answer as the Retort [20], as illustrated in (2).

(2) Speaker: Does Ed McMahon drink?
Addressee: Is the Pope Catholic? [20]

RQ+s, Retorts, and ISQs make the addressee search for a specific member within the domain of individuals, but RQ-s convey that the answer is the complement of the entire domain, which equals an empty set (being non-specific). RQ+s, Retorts and RQ-s appeal to the common ground (i.e., the shared knowledge between speakers [21, 5]), but not ISQs: the speaker of a RQ already has a certain answer in mind (and expects that the addressee has the same). ISQs do not appeal to the common ground, as the speaker does not have anyone in mind when asking a genuine question. Specificity and appeal to the common ground feature the four question types as in Figure 1:

Figure 1: Features of question meaning.



1.2. The prosody of RQs

It is commonly assumed that ISQs and RQs are prosodically distinct, but empirical studies on the

prosody of RQs are still rare and limited to RQ-s in English, Icelandic, and German [1, 23, 6, 16]. In particular, the distribution of boundary tones and nuclear pitch accent types, duration of utterances, and voice quality are found to contribute to the difference between ISQs and RQ-s, although these cues are weighed differently across the three languages. Taken together, these studies support the assumption that ISQs and RQ-s can be distinguished prosodically.

However, it is as yet unknown whether there is also prosodic difference among the subtypes of RQs (i.e., RQ- vs. RQ+ vs. Retort) and whether there exists prosodic distinction between ISQs and RQs in tone languages, such as Cantonese in our case.

1.3. Intonation in Cantonese

As a tone language, the F0 contour of Cantonese sentences necessarily involves the interaction between lexical tone and intonation. Overall, however, the intonation patterns of Cantonese are similar to those of many languages [4, 18]: statements generally have a falling F0 contour, and questions have a rising F0 contour, regardless of the canonical form of the tone [13], indicating the presence of a boundary tone.

Most studies on Cantonese question intonation focus on intonation questions (i.e., a declarative sentence with a rising contour), contrasting them with declaratives [13]. Cantonese sentence-final particles (SFPs) express similar functions as intonation in intonation languages, conveying the speaker's attitude or emotions. Studies on the intonation of questions suffixed by SFPs, however, are relatively rare [25, 24, 26], and even in those studies, the F0 contour of SFPs themselves is seldom examined. Given that SFPs are abundant in the language and that utterances in Spoken Cantonese typically end in SFPs [12, 22], this study focuses on the acoustic properties associated with two common SFPs *aa1* and *aa3*.

1.4. The Cantonese SFPs *aa1* and *aa3*

The particles *aa1* 吖 and *aa3* 呀 are chosen because of their high frequency of use in a wide range of contexts. Both ISQs and RQs are reported to feature *aa1* and *aa3*: *aa1* is claimed to make an utterance more emotive and possibly indicates a contrast, while *aa3* is described as "highlighting the relevance of the utterance" [14, 11, 22].

To analyze the prosodic properties of the SFPs *aa1* and *aa3* in different question types in Cantonese, we conducted a production experiment. The experiment is designed to examine how question types impact the prosodic realization of the two SFPs in *wh*-questions. The present study therefore contributes to this line of research in two directions by (i) examining more subtypes of RQs and (ii) extending to a tone language.

2. EXPERIMENT

2.1. Stimuli

We created 12 target *wh*-questions that are ambiguous between rhetorical and information-seeking readings, have the same syntactic structure and contain the same number of syllables. The *wh*-word in all target sentences is *bin1go3* 'who'; half of the target sentences end with the SFP *aa1*, and the other half end with the SFP *aa3*. To counter the effect of tonal co-articulation, the lexical tone of the syllable directly preceding the SFP is also controlled for and is balanced between the two SFPs. Cantonese has six unchecked tones, thus one sixth of the target sentences has Tone 1 before the SFP, one sixth has Tone 2, and so on.

For each of these 12 target sentences, we generated four short contexts, each favoring one of the four readings respectively (See Table 1: contexts are translated from Cantonese). The context descriptions are as concise and informative as necessary. The target sentence is placed in the final position within each context. In addition, each target sentence is preceded by another short sentence that is semantically congruent with the target sentence. The purpose of these additional sentences is to strengthen the intended reading. The register of both contexts and target sentences is deliberately chosen to reflect the spoken language.

Two experiment lists were created that differ in the combination of SFPs and target sentences: target sentences ending with *aa1* on the first list are replaced with *aa3* on the second list, and vice versa. The items were randomized within each list, but the same random order was used for the participants assigned to the same list. Additional 48 filler context descriptions were used with assertions as target sentences. With three practice trials, the total number of items was 99, which took an hour for the participants to complete on average.

2.2. Participants

We tested 21 speakers, but discarded the data from seven speakers due to difficulty in reading spoken Cantonese, cell phone use during the experiment, persistently skipping contexts, and extremely unnatural production. The remaining 14 participants are all native speakers of Cantonese (mean age: 19.6

Table 1: The four contexts for the target sentence *Jau5 bin1go3 soeng2 jam2 gaa1fe1 aa1/aa3?* have who like drink coffee SFP 'Who wants to drink coffee?'

ISQ: You are having a family gathering in your home with a lot of people. After lunch, you want to serve tea and coffee, but you don't know how much to prepare. You prepare the coffee first, and you want to find out how many cups are needed. So you ask:
"I don't know how many people want coffee. **Who wants to drink coffee?**"
RQ-: Mary is throwing a party. It's 1am and everybody is about to leave, but she is wondering if she should still make some coffee. You think nobody would drink any — some of them have already left anyway. So when Mary is telling you that she'll start the coffee machine, you say:
"Mary, wait, you shouldn't start the machine. **Who wants to drink coffee?**"
RQ+: Mary and you are roommates, and you had a guest from South America who brought you five different South American coffees. But neither you nor Mary drinks coffee, so you suggest to give all the coffee to someone. You just need to find a coffee drinker. The first person that comes in mind to both of you is John, your neighbor, because you both see him with his coffee on the balcony every morning, afternoon and evening — John is a real coffee-holic! So when Mary wonders who you should give all this coffee, you point to John, who's drinking coffee on his balcony right now, and say:
"Just look at that balcony. **Who wants to drink coffee?**"
Retort: The coffee machine went wrong. John is a real coffee lover; he drinks a lot of coffee every day. When you and John talk about the coffee machine, he asks you who should fix the coffee machine. You suggest that it should be him because he's the one who's desperate about getting coffee. He asks, "Who should fix the coffee machine?" You say:
"Who should fix the coffee machine? **Who wants to drink coffee?**"

years; *SD* = 1.2; 10 females, 4 males). They are all undergraduate students recruited via participant pools and are compensated by getting partial course credit for their participation in the experiment.

2.3. Procedure

The experiment was conducted in a sound-attenuated booth. In each trial, the context was presented auditorily (as recorded by a female native speaker of Cantonese) through a headphone and in written form on the screen simultaneously. The target sentence, along with the facilitating sentence, then appeared on the same screen. Participants were tasked to read both sentences exactly as written and as naturally as possible. Participants pressed a button to proceed to the next trial.

2.4. Data analysis and results

Each target sentence was extracted from the recordings and checked manually. A target sentence was excluded from the subsequent analysis if (i) the participant skipped the context for that trial, (ii) there was disfluency in the sentence, (iii) the SFP was omitted, and (iv) the participant made mistakes (e.g., by adding extra words or ignoring some words). In addition, one experiment item had to be removed because it contained one extra syllable in the target sentence. Accordingly, 565 target sentences entered the analysis. The distribution of them among question types and SFPs is summarized in Table 2.

Table 2: Distribution of target sentences included in the analysis.

	ISQ	RQ-	RQ+	Retort
aa1	70	71	69	72
aa3	73	68	69	73

The SFPs in all target sentences were annotated with Praat [2] by the authors. The durations of the two SFPs were extracted automatically. The F0 of the SFPs was estimated using Praat's autocorrelation periodicity detection algorithm[1] and manually checked. For the SFP in each target sentence, ten equidistant points were then extracted after pitch smoothing with a bandwidth of 20 Hz. We excluded SFP tokens produced with creaky voice (83 out of 565) from the statistical model for F0 contour, as the resulting irregularity in F0-tracking might obscure the intonation patterns.

For the statistical analysis of F0 contour and duration, we ran a set of mixed effects models. In order to characterize F0 contour, a least-squares linear regression was fit for each token, with the log-transformed F0 values from the ten equidistant points as the response and the corresponding time-offsets from the beginning of the syllable as the predictor. The slopes of these regression lines were then used as the dependent variable for the mixed effects model. For duration, the log-transformed values were used as the dependent variable for the model.

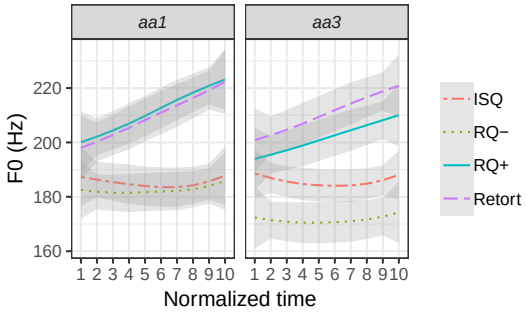
The fixed effects for all models included Question Type (ISQ vs. RQ- vs. RQ+ vs. Retort) and Particle (*aa1* vs. *aa3*), as well as their two-way interactions. The random effects structure was as maximally specified as possible: Participant and Item were included as random effects for the models of F0 contour, but only Participant was specified as the random effect for the models of duration (because including Item resulted in singular fit). The structure included only

a by-Participant random slope for Particle, as the models with more complex random structures failed to converge. P-values were calculated using the Satterthwaite approximation of degrees of freedom.

2.4.1. F0 contour

The time-normalized F0 contours of the two SFPs across question types are depicted in Figure 2. Significant fixed effects were found for Question Type of RQ+ vs. ISQ ($\beta = 9.2\times10^{-4}$, $SE = 2.9\times10^{-4}$, $p < 0.01$) and for Retort vs. ISQ ($\beta = 9.4\times10^{-4}$, $SE = 2.9\times10^{-4}$, $p < 0.01$), indicating that the rising F0 contour in RQ+ and Retort contexts is steeper than that in ISQ contexts. However, the F0 contours between RQ- and ISQ ($\beta = 7.9\times10^{-5}$, $SE = 2.9\times10^{-4}$, $p = 0.79$), and between RQ+ and Retort ($\beta = 2.2\times10^{-5}$, $SE = 2.9\times10^{-4}$, $p = 0.94$) were not significantly different. The fixed effects of Particle of *aa1* vs. *aa3* and all the interaction terms were not significant, suggesting the SFPs behave similarly.

Figure 2: F0 contours of the SFPs across question types. The contours are smoothed with LOESS, and the ribbons display 95% confidence intervals.



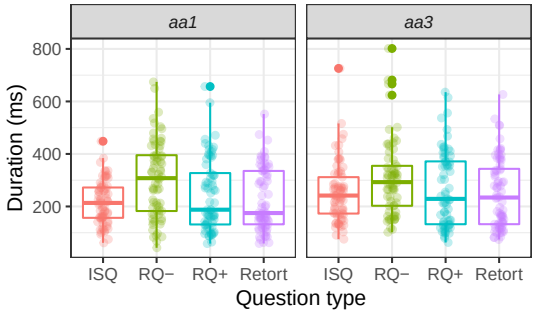
2.4.2. Duration

The distribution of the durations of the two SFPs across question types is shown in Figure 3. The only significant effect was for Question Type of RQ- vs. ISQ ($\beta = 0.25$, $SE = 0.071$, $p < 0.01$), indicating that the duration of SFPs in RQ- contexts is longer than that in ISQ contexts. The durations of SFPs among ISQ, RQ+, and Retort contexts were not significantly different. All the other effects and interactions were not significant.

3. DISCUSSION AND CONCLUSION

The results from our production study show that the SFPs *aa1*/*aa3* are realized distinctly in terms of F0 contour and duration under different question types.

Figure 3: Distribution of the durations of the SFPs across question types.



RQ-s are found to have longer SFP duration than ISQs, RQ+s and Retorts. The SFPs in Retorts pattern with RQ+s, both displaying a higher rising F0 contour, compared to the RQ- and ISQ conditions.

The three-way distinction in the suprasegmental properties of these questions is expected, considering the semantic differences between them. The different patterning of RQ+s and Retorts versus ISQs and RQ-s could be attributed to differences in affective stance [17]. RQ+s and Retorts are uttered in a context where the speaker believes that there is a specific salient answer to the uttered question, one that the addressee should be aware of. These utterances come with an attitude expressing 'obviousness' (or 'you should have known'). The rise in F0 of RQ+s and Retorts could thus potentially be attributed to engaging the addressee, calling upon their world knowledge.

Some of our findings tie in with what has been found for the prosody of ISQs and RQs in other languages. For instance, our finding that the SFPs in RQ-s do not show a rising F0 contour agrees with the finding that RQ-s tend to have more low boundary tones in English and German [23, 9]. Additionally, our result that the SFPs in RQ-s are longer, in comparison with those in ISQs, also echoes the finding that the sentence-final object is longer for RQ-s than for ISQs in German and Japanese [23, 15].

While the literature on the interface of meaning and intonation suggests a divide in the final contour of assertive and inquisitive speech acts [18, 7], our results show that RQ+s pose a challenge to this picture. On this view, RQ+s are expected to behave the same way as RQ-s. Our view on the different question types can account for the three-way prosodic differences.

We are currently investigating whether there is a global F0 and durational difference over the entire utterance among different question types and devising a perception experiment.

4. REFERENCES

[1] Banuazizi, A., Creswell, C. 1999. Is that a real question? Final rises, final falls and discourse function in *yes-no* question intonation. *Proc. CLS 35* Chicago, IL. 1–14.

[2] Boersma, P., Weenink, D. 2018. Praat: Doing phonetics by computer [computer program].

[3] Caponigro, I., Sprouse, J. 2007. Rhetorical questions as questions. *Proc. of Sinn und Bedeutung* Barcelona. 121–133.

[4] Chan, Y.-Y. F. 1974. *A Perceptual Study of Tones in Cantonese*. Hong Kong: Centre of Asian Studies, University of Hong Kong.

[5] Clark, H. H. 1996. *Using Language*. Cambridge: Cambridge University Press.

[6] Dehé, N., Braun, B., Wochner, D. 2018. The prosody of rhetorical vs. information-seeking questions in Icelandic. *Proc. the 9th International Conference on Speech Prosody 2018* Poznań. 403–407.

[7] Farkas, D. F., Roelofsen, F. 2017. Division of labor in the interpretation of declaratives and interrogatives. *Journal of Semantics* 34, 237–289.

[8] Han, C.-H. 2002. Interpreting interrogatives as rhetorical questions. *Lingua* 112, 201–229.

[9] van Heuven, V. J., Haan, J. 2000. Phonetic correlates of statement versus question intonation in Dutch. In: Botinis, A., (ed), *Intonation: Analysis, Modelling and Technology*. Dordrecht: Springer 119–143.

[10] Krifka, M. 2011. Questions. In: Maienborn, C., von Heusinger, K., Portner, P. H., (eds), *Semantics: An International Handbook of Natural Language Meaning* volume 1. Berlin: De Gruyter Mouton 1742–1785.

[11] Li, B. 2006. *Chinese Final Particles and the Syntax of the Periphery*. PhD thesis Leiden University Leiden.

[12] Li, C. N., Thompson, S. A. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press.

[13] Ma, J. K., Ciocca, V., Whitehill, T. L. 2006. Effect of intonation on Cantonese lexical tones. *J. Acoust. Soc. Am.* 120(6), 3978.

[14] Matthews, S., Yip, V. 1994. *Cantonese: A Comprehensive Grammar*. London: Routledge.

[15] Miura, I., Hara, N. 1995. Production and perception of rhetorical questions in Osaka Japanese. *Journal of Phonetics* 23, 291–303.

[16] Neitsch, J., Braun, B., Dehé, N. 2018. The role of prosody for the interpretation of rhetorical questions in German. *Proc. the 9th International Conference on Speech Prosody 2018* Poznań. 192–196.

[17] Ochs, E., Schieffelin, B. 1989. Language has a heart. *Text* 9, 7–25.

[18] Pierrehumbert, J., Hirschberg, J. 1990. The meaning of intonational contours in discourse. In: Cohen, P. R., Morgan, J., Pollack, M. E., (eds), *Intentions in Communication*. Cambridge, MA: MIT Press 271–312.

[19] Rohde, H. 2006. Rhetorical questions as redundant interrogatives. *San Diego Linguistics Papers* 2, 134–168.

[20] Schaffer, D. 2005. Can rhetorical questions function as retorts? Is the Pope Catholic? *Journal of Pragmatics* 37, 433–460.

[21] Stalnaker, R. 1978. Assertion. *Syntax and Semantics* 9, 315–332.

[22] Sybesma, R., Li, B. 2007. The dissection and structural mapping of Cantonese sentence final particles. *Lingua* 117, 1739–1783.

[23] Wochner, D., Schlegel, J., Dehé, N., Braun, B. 2015. The prosodic marking of rhetorical questions in German. *Interspeech 2015: 16th Annual Conference of the International Speech Communication Association* Dresden. 987–991.

[24] Wu, W. L. 2009. Sentence-final particles in Hong Kong Cantonese: Are they tonal or intonational? *Proc. INTERSPEECH 2009 - 10th Annual Conference of the International Speech Communication Association* Brighton. 2291–2294.

[25] Xu, B. R., Mok, P. 2011. Final rising and global raising in Cantonese intonation. *Proc. 17th ICPhS* Hong Kong.

[26] Zhang, L. 2014. Segmentless sentence-final particles in Cantonese: An experimental study. *Studies in Chinese Linguistics* 35, 47–60.

[1] The setting for each parameter of the algorithm was as follows: time step = 0.001 s, maximum number of candidates = 15, silence threshold = 0.03, voicing threshold = 0.45, octave cost = 0.01, octave-jump cost = 0.35, voiced/unvoiced cost = 0.14. The values for pitch floor and pitch ceiling were adjusted based on the F0 range of the target sentence in question. The value for the time step was set to be smaller than the default so that the algorithm returned more pitch values. The other values were all algorithm defaults.