

Prosody and Dialogue Act: A Perceptual Study on Chinese Interrogatives

Gan Huang^{1†}, Aijun Li^{1,2*}, Sichen Zhang², Liang Zhang³

¹Institute of Linguistics, Chinese Academy of Social Sciences, China

²University of Chinese Academy of Social Sciences, China

³China University of Political Science and Law, China

†huanggan666@yeah.net, *liaj@cass.org.cn

Abstract

Prosody conveys dialogue acts and intentions in speech interaction. This study aims at investigating the interplay between prosody and dialogue acts (pragmatic functions) for Chinese dialogues. To this end, a perceptual experiment was carried out on interrogative intonations with varied prosodic features and contexts associated with 3 dialogue acts including *request for affirmation*, *backchannel*, and *elaboration*. The results demonstrated that (1) the dialogue acts of the context affects the perception of the interrogatives with the same prosodic features; (2) when there is a mismatch between the actual prosody and the context, the context limits or reduces the perception gradient of interrogatives; (3) the perception of interrogative information mainly depends on prosody, while context make contributions to the perception of interrogative-declarative information by modulating listeners' interpretation of dialogue acts performed by certain prosodic features.

Index Terms prosody, context, dialogue act, pragmatic function, interrogative perception

1. Introduction

Interrogative information coding has always been a hot topic in linguistics, natural speech processing and other related fields. Many problems have been solved, yet more provoked.

Like syntactic markers (question words, modal particle, and the change of word order), phonetic markers (intonation) are commonly used by speakers to express interrogative information. Quirk et al. [1] pointed out that in English, sentences of different syntactic structures and different semantic functions have different intonation patterns. This relationship between intonation and semantic functions has been proved by quantities studies, for instance, the prosody of English marks the boundary of linguistic units [2-6], as well as the interrogative-declarative information of the sentence [2, 7-9]. Several studies on Mandarin also revealed that the overall pitch of Mandarin interrogative sentences is generally higher than that of declarative sentences; meanwhile, the perception of interrogatives primarily relies on the pitch range, duration and pitch slope of the boundary syllable [10-12].

Recent focus of research has been shifted from the relationship between intonation and declarative-interrogative sentences to the interplay between prosody and context.

For starters, context, considered as more than the text preceding and following target speech, also serves to carry interrogative information. The actual situation of the current conversation indicates the role played by each utterance, that is, its pragmatic function or, as specified in the present study, dialogue act. Within a specific utterance, the context is

constructed by both prosody and dialogue act, and the association in between provides evidence for speech perception [13, 14]. Liu and Li [15] found that compared to prepared speech, spontaneous speech display greater variation in the prosodic features of pitch, duration and pitch contour. Li et al. [16, 17] observed that apart from interrogative sentences marked by lexical markers, quite a few declarative questions, for example, echo questions, are not clearly distinguished from statements in prosody. That is to say, based on intonational phonology theories and laboratory experience, echo questions in isolation will be perceived as statements. In other words, context modulates the threshold of speech mood perception.

Research has further established the correlation between certain prosodic features and pragmatic functions. Gravano et al. [18] investigated the intonation of the affirmative word "okay" in English. They found that the same words with different boundary tones and word-level intensities convey different dialogue acts. In Li et al.'s [17] study of Mandarin echo utterances, speakers can apply different prosodic features, such as pitch slope, tone register, speech rate, etc., to realize the dialogue acts corresponding to given contexts, including request for affirmation (C1), backchannel (C2), elaboration (C3), and request for explanation (C'). As shown in Figure 1, in terms of mean F0 values, C' is noticeably higher than C1 and C3, while C1 is only slightly higher than C3. C2 appears to be the lowest. The results of Li et al.'s [17] classification analysis reveal that these prosodic features contribute to classify interrogative and declarative sentences, and further classify dialogue acts.

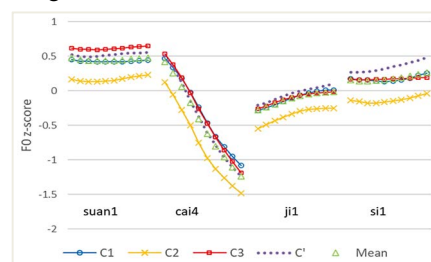


Figure 1: Pitch pattern of *suan1cai4ji1si1* (酸菜鸡丝) under different dialogue acts. [17]

Therefore, there is an interplay between prosody and dialogue acts in the encoding and decoding process of interrogative information, demonstrated in Figure 2. For one thing, prosodic features, bearing illocutionary force to realize dialogue acts, provide evidence for listeners' perception of interrogative information. Since an utterance classified as an interrogative sentence performs the dialogue act of asking a question, listeners rely on their concept of phonological-

semantic intonation prototype to make judgments of whether the present utterance approximates to typical interrogatives. For another, when the prosodic features show less similarities to any phonological-semantic intonation prototype, the modulation effect of context plays in. Listeners' concept of contextual-pragmatic intonation prototype, including that of pre- and post-utterance context such as the previous dialogue and response, contributes to their explanation to the dialogue act performed by certain prosodic features.

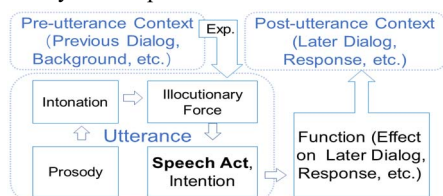


Figure 2: Relation of prosody, context and dialogue acts.

The present research focuses on this interplay between prosody and dialogue acts in encoding and decoding of interrogative sentences in Mandarin. To elaborate, this research aims at answering the following questions: (1) would the manipulation of context, varying interlocutors' responses to the questions, influence the judgement of dialogue acts with the same prosody; (2) how would the listener perceive an utterance when there is a mismatch between the actual prosody and the prosody constrained by context; (3) in different contexts, how would prosodic changes influence the perception of interrogatives?

2. Experiments

2.1. Stimuli

Table 1: Variables in the study.

Condition	Variable
Context	Isolation C0
	Followed by an affirmative answer C1
	followed by indirect response C2
	Followed by Evaluative response C3
Dialogue Act	Echo to ask theoretically related to C1
	Echo as backchannel theoretically related to C2
	Echo to elaborate theoretically related to C3
Prosody (Figure 3)	Boundary tone target (3a) BPs (0st, +2st, +4st)
	Boundary tone register (3b) BLV (-2st, 0st, +2st)
	Sentence register (3c) SLV (-1st, 0st)
	Preceding Boundary tone duration (3d) BDr (-25%, 0, +25%)
	Boundary tone duration (3d) PDr (-25%, +25%)

*st = semi-tone

2.1.1. Manipulation of prosody

A speech of interrogative phrase *suan1cai4ji1si1* (酸菜鸡丝, shredded chicken) in Context 1 (C1) was selected as the standard sample from the recording of a 24-year-old female speaker in Li et al.'s experiment [17]. With the reference to

[10-12, 17], the prosodic features listed in Table 1 were manipulated as illustrated in Figure 3. Since the speakers tend to produce C2 with a pattern of lengthening the entire utterance and lowering the overall pitch, the changes of the boundary tone appears to be the most prominent in the context. Hence, there were 3 manipulated stimuli of each boundary tone parameters, 5 of each duration parameters, and 2 of the sentence pitch parameters. Moreover, since the target sentences are acoustically and perceptually similar to declarative sentences, only the target value of boundary tones is manipulated. The specific features and methods of manipulation is shown in Table 1 and Figure 3. These stimuli were obtained through manually adjusting the target values of boundary tones followed by changing other variables via self-developed script in Praat 6.1.3. In this way, we obtained 90 versions of the speech samples.

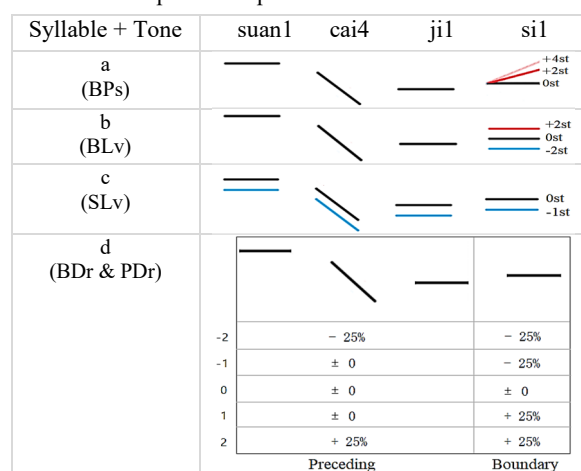


Figure 3: Scheme of prosody manipulation.

2.1.2. Manipulation of context

Since the textual information of the target sentence is limited, the interlocutor's responses to the target sentence provide the context information that has significant influence on the listener's judgments of dialogue acts.

As shown in Table 2, the dialogue acts of the target sentence were manipulated by changing the responses of the listeners. The key parts of the dialogue are listed in Table 2. The 90 versions of target sentences were inserted into 3 dialogue contexts, making 270 dialogues in total.

Table 2: Target sentence in 4 contexts.

Dialogue acts	Dialogues
C0	酸菜鸡丝 A shredded chicken.
C1 To ask	M: 点个酸菜鸡丝吧 F: 酸菜鸡丝 M: 对 M: I'll order a shredded chicken F: (.2s) A shredded chicken M: (.2s) Yes
C2 Backchannel	M: 再来个酸菜鸡丝 F: 酸菜鸡丝 F: 还需要别的吗 M: (1.4s) And a shredded chicken F: (.3) A shredded chicken F: (1.4s) Anything else
C3 To elaborate	M: 他点的什么菜 F: 酸菜鸡丝 M: 就一个啊 M: What did he order F: (1s) A shredded chicken M: (.3s) Only one

2.2. Procedure

Two experiments were designed to examine the interplay of prosody, context and dialogue acts, respectively in isolation (C0) and in dialogue contexts (C1, C2, C3). The perception experiments were conducted on the platform E-Prime 2.0.

2.2.1. *Exp. 1: Isolated stimuli*

Two laptops were applied to conduct the experiment in turn and the sound output was controlled by the same model of sound card. In this experiment, the subjects were asked to react to 90 pseudo-random arranged stimuli. The text showed up as white on black on the monitor 3000ms before the sound signal. A 7000ms blank was inserted between two trails. The subjects can press number 0 to replay the sound during the process. The subjects were required to evaluate the possibility of the stimulus being a question by choosing number 1-5 (1 for 0%, 5 for 100%).

10 gender-balanced Mandarin speakers (mean age 24 years) without hearing and speaking impairment participated in the experiment. None of them was involved in Exp. 2.

2.2.2. *Exp. 2: context stimuli*

We applied a 3*3 Latin square design in the 270 dialogues. Similar tasks and procedures with Exp. 1 were applied in this experiment. Target sentences showed on the monitor were bold and yellow. An 12000ms blank was inserted between two trails. The subjects can press number 0 to replay the sound.

30 gender-balanced Mandarin speakers (mean age 22.5 years) without hearing and speaking impairment were divided into 3 groups in the experiment.

3. Analysis and Results

3.1. The effect of context manipulation

In this part, we examine the influence of contextual manipulation in the perception of dialogue acts with the same prosody, so as to answer the first research question.

In Figure 4, the horizontal axis is the serial number of the stimuli in an ascending sort order of the mean possibility evaluating score, and the vertical axis is the mean score of the subjects' evaluation. The higher the mean score, the higher the propensity of the listener to perceive the stimulus as an interrogative sentence.

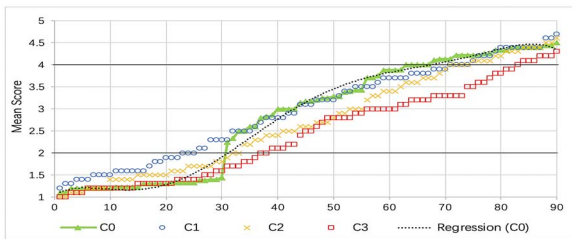


Figure 4: Scores of stimuli in 4 contexts.

The mean score of C0 is regarded as the reference pattern. Changes of slope could be seen in different score ranges (<2, 2-4, >4), which further rationalized the classification of typical statements, typical questions and the utterances in between.

In typical statements (mean score <2), the target sentence in C1 is followed by an affirmative answer. It carries more interrogative information, which increases the overall score compared to C0. In C3, since the context corresponds with the prosody, the mean score pattern copies that of C0. The context in C2 is neutral, so does its effect on the scores. Hence, the mean score lies in between the curves of C1 and C3.

For utterances scored from 2-4, as interrogative features in prosody intensify, C1 and C0 demonstrate very similar

patterns, which means the boost effect of C1 is weakened by the manipulation of prosody. However, C3 consistently shows an obvious contextual effect on the listeners' perception, with the scores for each stimulus much lower than those for C0 and C1. The curve of C2 still lies between that of C1 and C3.

In typical questions (score >4), the effect of neutral context on C2 yields to effect of prosodic features. Hence, C2 shows a similar pattern to C1 and C0. C3 still shows a clear effect on the scores.

The results support the findings in Li et al. [16] that the context influences the perception of interrogative information, which is manifested on the deviation in comparison to C0. However, to which degrees the context could modulate is due to the manipulation of the prosodic features. When changing a statement (<2) into a non-statement (>2), the context has significant effects: C1 (question) changes at an early stage, C3 (statement) at the last, and C2 (neutral) and C0 (isolation) in between. As for changing a non-question (<4) into a question (>4), though its effect is weakened by prosodic features, the context also functions in C3.

3.2. The mismatch between the prosody constrained by context and the actual prosody

In this part, we investigate the effect of the mismatch between the prosody constrained by context and the actual prosody, so as to answer the second research question. Figure 5 demonstrates the number of stimuli scored as typical statement (<2) and typical question (>4) in four contexts. Again, the results of C0 is regarded as the reference. Theoretically, the sentences in C1 are questions, with the prosody corresponding to the phonological-semantic intonation prototype of typical interrogatives, while those in C3 are statements, with the prosody corresponding to typical declaratives.

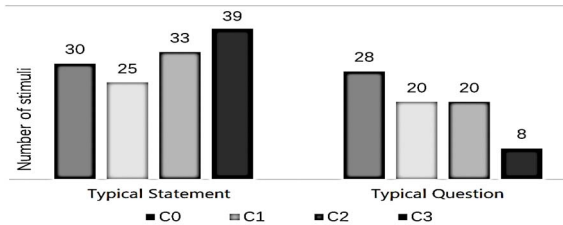


Figure 5: Numbers of stimuli as typical statement/question in different contexts.

Statements in C1 (an interrogative context) and questions in C3 (a declarative context) undergo mismatch between their actual prosody and the prosody constrained by context. The data illustrates when there is a mismatch, the context perform a modulation effect of reducing the influence of prosody on the perception. Compared to C0, the number of typical statements in C1 declines from 30 to 25, and the number of typical questions in C3 also declines from 28 to 8.

3.3. Interplay of prosody and context

The General Linear Model (GLM) is adopted to analyze the effects of the prosodic parameters and different contexts.

The results show the main effect of all the manipulations in interrogative perception (p<0.001), except boundary tone register (BLV, p>0.05). The increased boundary tonal target and sentence register boost the possibility of the stimulus to be perceived as an interrogative sentence. The shortened durations of the preceding boundary and the current

boundary, or the current boundary only, decrease the possibility. This possibility is also significantly influenced by the manipulation of context: the scores go higher as the context requires the target sentence to be a question.

The prosodic parameters and the contextual background influence the scores interactively, as shown in the two figures below. In Figure 6, the effect of strengthening the interrogatives by increasing boundary tone target is modified by different contexts ($\eta^2=0.012$, $p<0.001$). Without the dialogue context, like in C0, the subjects are more sensitive to the change of boundary tone pitch target, comparing to the other contexts. In Figure 7, the effect of strengthening the interrogatives by an increased sentence register ($\eta^2=0.003$, $p<0.05$) is also modified by different contexts. C0 reflects a weaker effect of the register change, so does C3. C1 and C2 present a similar effect of strengthening the interrogative mood by a comparatively higher sentence level register.

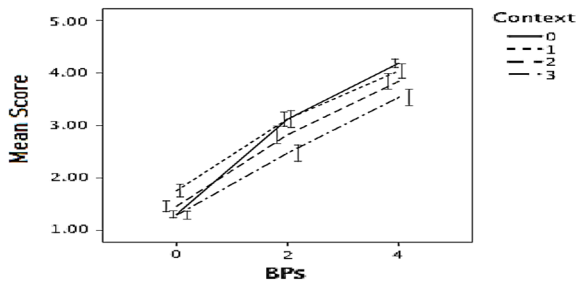


Figure 6: Effect of boundary tonal target and context.

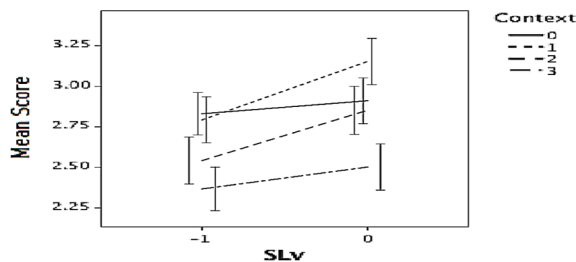


Figure 7: Effect of sentence register and context.

Hence, a significant interplay is found among speech duration, boundary tone registers and context ($\eta^2=0.012$, $p<0.001$), as shown in Figure 8. The lengthened preceding boundary tone duration and boundary tone duration increase the possibility of stimuli to be perceived as interrogative sentences. However, the degree of interrogation is modulated by the boundary register and the context.

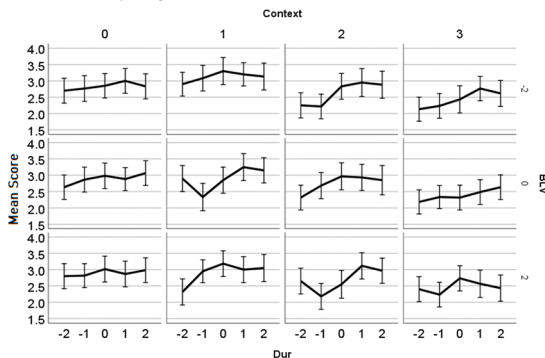


Figure 8: Effect of duration, boundary register and context.

4. Discussion

In this perceptual study, a group of Mandarin speakers were asked to score the possibility of an utterance being an interrogative sentence in different contexts with different prosodic features.

The effect of context on the scoring supports the findings in Li et al. [16]. That is, context influences the perception of interrogative information, but in varied degrees as a consequence of different prosodic features.

When there is a prosody-context mismatch, the context consistently limits or reduces the effect of prosody. When the prosody matches the contextual requirement, the pragmatic functions diverged. To be specific, compared with C0, C3 increases the number of stimuli scored as a typical statement from 30 to 39. Thus, we predict that a similar pattern would appear in C1, with the same number of stimuli (if not increased) scored as a typical interrogative sentence, i.e. C0. However, the number decreases from 28 to 20. A close look was paid to the number of stimuli scored between 3 and 4. The number of stimuli in C1 (27) is higher than that in C0 (19), and the total numbers of stimuli scored higher than 3 are balanced in C1(47) and C0 (47). This suggests that in the spontaneous speech, with the help of context, speakers may not use a prosodically-typical interrogative sentence to request an answer. Therefore, the phonological-semantic prototype of interrogative sentence in C1 may not be scored as a 4 or 5 typical question, indicating that there might be a larger contextual-pragmatic prototype in perceiving interrogative sentences in C1.

Furthermore, the manipulation of prosodic features greatly influences the scores in all types of contexts. The boundary tone register plays a prominent role, leading to a sharp increase in the C0 perception (scored from 1.44 to 2.25).

Finally, the interplay between prosodic and contextual manipulations is also examined. The increased tonal target and the increased pitch on the sentence level raise the listener's tendency to perceive the discourse as interrogative, which is accord with Lin [11, 12]. Meanwhile, the context also modulates the effect of boundary tone and sentence register. C0 is more sensitive to tonal target change, but less sensitive to sentence register change.

5. Conclusion

This study investigated the interplay effects of prosody and dialogue acts in the perception of interrogative intonation. The results prove and clarify the following interplay effects: (1) the dialogue acts of the context affects the perception of the interrogatives with the same prosodic features; (2) when there is a mismatch between the actual prosody and the context, the context limits or reduces the perception gradient of interrogatives; (3) the perception of interrogative information mainly depends on prosody, while context make contributions to the perception of interrogative-declarative information by modulating listeners' interpretation of dialogue acts performed by certain prosodic features.

6. Acknowledgements

This work is supported by the Key NSSFC (No. 15ZDB103), the National Key R&D Program of China (2017YFE0111900, 2013CB329301), and the Innovation Program of Chinese Academy of Social Science.

7. References

- [1] R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik, *A comprehensive grammar of the English language*. London & New York: Longman, 1985.
- [2] J. Pierrehumbert, "The Phonology and Phonetics of English Intonation," Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1980.
- [3] M. E. Beckman and J. B. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology*, vol. 3, pp. 255–309, 1986.
- [4] A. Cruttenden, *Intonation*. Cambridge: Cambridge University Press, 1986/1997.
- [5] C. Gussenhoven, *The Phonology of tone and Intonation*. Cambridge: Cambridge University Press, 2004.
- [6] D. Ladd, *Intonational phonology*, (2nd ed.). Cambridge and New York, NY: Cambridge University Press, 2008.
- [7] M. Halliday, *Intonation and Grammar in British English*. The Hague: Mouton, 1967.
- [8] D. Bolinger, "Intonation and its parts," *Language*, vol. 58, pp. 505–533, 1982.
- [9] J. t'Hart, R. Collier, and A. Cohen, *A Perceptual Study of Intonation: An Experimental-phonetic Approach*. Cambridge: Cambridge University Press, 1990.
- [10] J. Yuan, C. Shih, and G. Kochanski, "Comparison of declarative and interrogative intonation in Chinese," in *Proceedings of the 1st International Conference on Speech Prosody*, pp. 711–714, 2002.
- [11] M. Lin, "Interrogative vs. declarative and the boundary tone in Standard Chinese," *Studies of the Chinese Language*, no. 4, pp. 364–376, 2006
- [12] M. Lin, "Similarities and differences between Chinese and English intonation and the teaching of Chinese intonation to CFL learners: my view on how to avoid a foreign accent," *Journal of International Chinese Teaching*, no. 3, pp. 39–46, 2015
- [13] H. Sacks, E. Schegloff, and G. Jefferson. "A simplest systematics for the organization of turn-taking for conversation". *Language*, 50(4): 696–735, 1974.
- [14] E. Schegloff, *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press, 2007.
- [15] Y. Liu, and A. Li, "Comparative analysis between read and spontaneous speech," *Journal of Chinese Information Processing*, vol. 16, no. 1, pp. 13–18, 53, 2002.
- [16] A. Li, Y. Jia, X. Liu and L. Zhang, "Decoding of echo question in Chinese spoken dialogue," in *ICPLC 2013 – International Conference on Phonetics of the Languages in China*, December 2–4, Hongkong, China, 2013.
- [17] A. Li, G. Huang, and Z. Li, "Prosodic cues in the interpretation of echo questions in Chinese spoken dialogues," in *APSIPA ASC 2019 – Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2019*, November 18–21, Lanzhou, China, 2019.
- [18] A. Gravano, J. Hirschberg, and S. Benus, "Affirmative cue words in task-oriented dialogue," *Computational Linguistics*, 38(1), pp. 1–39, 2012.

[This paper was published at 12th International Symposium on Chinese Spoken Language Processing (ISCSLP) 2021]