PQSR: A Speech Corpus of Polar Questions and Spontaneous Responses in Standard Chinese with Complex Intentions Annotated

Yingyi Luo, Yue Huang, Qingke Sun, and Shuwen Chen
Institute of Linguistics, Chinese Academy of Social Sciences
E-mail: luoyingyi@cass.org.cn Tel: +86-10-85195375

Abstract— **To advance research of how communicative intentions are conveyed in conversations and how prosody plays a role, we created PQSR, a context-specific, intention-annotated speech corpus of polar-questions and responses in Standard Chinese. It consists of two subsets: Polar Question Production (PQP) with 103 native speakers posing polar questions in given contexts for information seeking or rhetorical expression, and Spontaneous Responses (SR) with 106 speakers' natural oral reactions to these questions. Data was collected using a combination of experimental and naturalistic designs, with participants recording themselves on their smartphones. Beyond phonetic annotations, we introduced a framework for classifying communicative intentions, focusing on the respondent's concerns regarding the questioner, along with a corresponding annotation system.**

I. INTRODUCTION

A. Motivation

In recent years, the development of human-computer interaction has led to the growth of conversation corpora or datasets. The majority of these datasets are application-scenario-constrained or topic-centered discussions primarily focusing on information retrieval [1-3]. For instance, many corpora primarily aim to resolve the knowledge- or evaluation-based problem [4-6]. However, human-human daily interactions serve purposes far beyond this. Particularly, the intentions underlying spoken communication are intricate and multifaceted, where contextual and vocal cues play crucial roles in conveying meanings that often transcend the literal words spoken, imply their opinions, and fulfill communicative functions [7].

This complexity becomes even more pronounced in situations when the dialogue involves the release of negative attitudes or emotions. For instance, it is common in daily life that one tries to comfort the other with "positive" words but in fact escalates tensions. However, datasets capturing such dynamic and intent-laden conversations are scarce. One possible reason is the challenges associated with documenting natural, unscripted conversations in real-life interactions. On the other hand, data collected from filmed materials or pure laboratory settings often lack authenticity due to the small and biased sample of participants and the inherent performative aspect of these environments.

To address this gap, we aim to provide a dataset of Standard Chinese speech communication with complex intentions. Particularly, we focus on polar questions and responses in interactional contexts involving both neutral and negative events and employed a hybrid of experimental and naturalistic design to collect data.

B. Communicative intention

In traditional linguistics, Searle's theory of speech acts is a seminal work that focuses on predicting speakers' purposes, identifying speakers' goals trying to achieve via speech such as asserting, commanding, requesting, promising, and expressing emotions, which can often be "indirect" in relation to the literal words used [8]. Building on this, Ref. [9] proposed a classification of dialogue acts, including suggest, greet, bye, accept, and reject, among others. Targeted at Natural Language Processing (NLP), the Segmented Discourse Representation Theory (SDRT) focuses on inducing rhetorical relations to comprehend discourse structure, including explanation, elaboration, contrast, counterevidence, etc., which can be seen as intentions [7].

On the computational front, with language data, neurobiological recordings induced by comprehension, or multi-modal measurements, there is a keen interest in training classification models for intent-related characteristics. Influentially, Ref. [10] defined 3 polarities and 22 specific sentiment classes based on online text conversations, spanning basic emotions like joy and surprise, and social behaviors such as complaint, criticism, disappointment, and opposition. Ref. [11] categorized 20 fine-grained intents into expressing emotions and attitudes versus achieving goals, based on data from TV series. Recently, sarcasm and irony have gained

This research was supported by National Natural Science Foundation of China (NSFC 62276185) and by Key Laboratory of Linguistics, Chinese Academy of Social Sciences (Project #2024SYZH001).

979-8-3315-7200-8/25/$31.00 ©2025 IEEE

significant attention in research [12, 13]. While these classifications have their applicable scenarios, they may overlook the multidimensional nature of communicative intentions, leading to inadequacy for depicting atypical question-answering situations.

C. *Polar-questions and responses in conversation*

Questions can be used to seek information from the interlocutor (referred to as information-seeking questions, or ISQs) or to make assertions, express (negative) feelings, or convey rhetorical meaning (referred to as rhetorical questions, or RQs). For polar questions, the expected answer may be 'yes' or 'no' when they perform the directive speech act of requesting information [14]. However, RQs does not expect an answer since they have the feel of assertion and usually the literal answers are already known to both parties. For instance, 'Have you drunk water?' in Table 1B expresses Mum's dissatisfaction of Andy not drinking water, rather than requesting an answer of yes or no. In a corpus study, Ref. [15] finds that answers to RQs are overwhelmingly confirmations and backchannels, suggesting that these are reactions from the addressee. The present study adopts this concept and defines the oral reaction to the polar questions as "responses".

Tbl. 1 Examples of polar questions in context.

1A. **Neutral context-ISQ**	1B. **Negative context-RQ**
Seeing Andy doing his homework, his mother asks him concernedly:	Knowing Andy hasn't drunk the water, his mother asks angrily:
Have you drunk your water?	*Have you drunk your water?*

The paralinguistic differences between ISQs and RQs have attracted considerable attention because identical texts can be employed in either way depending on the non-textual cues. Specifically, phonetic studies across languages constantly observed acoustic distinctions of RQs and ISQs, yet a general prosodic marking pattern for RQs remains to be concluded [16-20], likely due to the diversity of rhetorical meanings and intents that were examined. By comparison, fewer studies have investigated responses to RQs, primarily based on literary works [21, 22]. Controversial classifications have been proposed in terms of responses' surface forms or the meanings.

D. *Data collection in a combination of experimental and naturalistic design and intention annotation*

To support research in these areas, this corpus collects speech samples of questions and responses, particularly focusing on the intention expression in RQs and their responses. The experimental design primarily features highly controlled, scripted daily-life contexts and polar questions, generating string-identical polar questions for both ISQs and RQs: they convey speakers' intentions to seek information within a neutral context and to express dissatisfaction within a context involving negative events.

Furthermore, two naturalistic approaches are employed: recording speakers' spontaneous oral responses to context-questions, enabling free intention expression based on personal understanding; and adopting an online method using participants' smartphones for sound recording to enhance ecological validity. This yields high-quality speech data comparable to lab settings with respect to f0 signals [7] and minimizes performative tones and any lab-induced interference on emotional state, simulating real-life smartphone messaging. A larger, diverse speaker sample is anticipated as well.

Thus, PQSR comprises two sub-datasets collected separately: Polar Question Production (PQP) and Spontaneous Responses (SR). PQP is data of speakers uttering designated polar-questions with two intents labeled: acquiring information and expressing dissatisfaction. SR is data of spontaneous speech responses to the polar-questions within the context. Importantly, we have established the Response Intention Annotation System (Table 2). Fourteen fine-grained intents are labeled, falling into 5 dimensions, indicating the respondent's concerns in each utterance, the questioner's literal/factual doubt, attitude, emotion, and/or commitment, as well as conversation extension. This system distinguishes itself from the pioneering ones in that: 1) it focuses on to what aspect(s) of the questioner (from explicit to implicit levels) the respondent reacts, rather than the respondent's own manifestation regardless of the questioner. For example, "Emotion" here refers to the concern for the questioner's emotion rather than expressing one's own emotion. 2) Previous coarse-grained indirect acts have been refined, such as "explanation" [23] which may correspond to different labels in this system. 3) Categories that were previously detailed, such as complaint, disappointment, etc. [10], have been de-emphasized due to insufficient distinction in the current context. 4) An utterance may possess attributes across multiple dimensions (labels are connected with a hyphen "-"), reflecting that interactional intents are multifaceted.

II. Data Collection

PQR and SR exhibit considerable overlap in their data collection methods. Therefore, we will introduce both datasets together when discussing the methodology.

A. *Participant*

Native Mandarin speakers were recruited for monetary reward through advertisements that placed no restrictions on variables such as age, gender, geographical region, educational background, or other demographic factors. At the current stage, a total of 233 native Mandarin speakers from two groups attended (116 participants for PQP and 117 participants for SR). We removed data of participants who either failed to follow instructions or achieved suboptimal recording quality. So far, PQP comprises data of 103 participants (24 males and 79

females, aged 18-60) and SR comprises data of 106 participants (31 males and 75 females, aged 18-60).

B. Materials

To induce relatively natural speech of questions and spontaneous responses from participants, daily-life contextual information and the polar question to start the dialogue were created in experimental design. Within the contexts, conversations take place in the matchmaking, the company, the school/classroom, and at home.

Two types of scenarios involving either neutral events or negative events are paired with the polar question, so that the question is a straightforward inquiry to obtain information produced by a character in one context (ISQ), and is used in rhetorical manner in another context to express one's dissatisfaction toward the other interlocutor, a third party, or an object (RQ) (see Table 1 for examples).

A total of 84 polar questions ending with the sentence-final particle ma ("吗") were generated, with diverse structure and length ranged from 4 to 10 characters. We avoid using first- and second-person pronouns as the sentence-initial subjects because rhetorical meanings may be preferred for such construction. With each question pairing with two scenarios, a total of 168 sets of context-question were generated, half being neutral-ISQs and half being negative-RQs. Responses to the questions are not scripted.

The 168 sets were split into four versions. Each version contained 21 neutral-ISQs and 21 negative-RQs. Each polar question will only appear once in one scenario in each version.

C. Platform and Procedures

The experiment utilized the online platform Wenjuanxing (https://www.wjx.cn). Participants scanned a QR code to access the experiment on their smartphones in a quiet setting. They were advised to use headphones for better audio quality but were not restricted in their choice of mobile devices or recording equipment. Participants first read the contextual information on the screen silently, then pressed and held the microphone icon to record their voice. They were encouraged to immerse themselves in the scenario and speak in Standard Chinese. Participants could re-record if they made mistakes or felt it was necessary.

For the data collection of PQP, participants acted as the questioner. Each was randomly assigned to one of the four versions, and the 42 sets of context-question were presented in a random order. They were required to comprehend the context, and then read aloud the polar question that was shown on the screen in bold, consistent with the context and conveying the intention with appropriate intonation. The production of polar question was simultaneously recorded by themselves. For the data collection of SR, participants acted as the respondent in the conversation, not the questioner. Each participant was assigned to one version (42 trials), silently read and comprehend the context and the question, and was required to make a spontaneous oral response to the question. To obtain the natural response, no specific requirements were set regarding its content, intonation or length.

It took approximately 15 minutes per participant to complete PQP recording, and 20-30 minutes to complete SR.

III. Data Annotation

Currently, 4159 questions and 4361 responses have been collected. First, one data file was summed for questions and responses, respectively. Each item (a recording) was documented with their literal contents in text, context type (neutral/negative), intent label (2 for PQP and 14 for SR, Table 3), scenario categories (matchmaking, office, school, home), sentence structure (4 types: SVO, SV, VO, OV), word count and the subject ID. Further, each item includes a speech data file with sampling rate of 44100HZ and a corresponding text data file from transcription. The speech data consists of three files: 1) a Sound.wav audio file; 2) a TextGrid file for storing manually annotated content; and 3) a PitchTier file for displaying the trend of pitch contour. To eliminate the individual differences of intensity, the peak intensity of all speech file was adjusted to 0.99 using the standardization method of Scale Peak in Praat [23]. Each PitchTier was initially tracked by Praat and subsequently manually corrected.

A. Polar Question Production (PQP)

To advance the phonetic and phonological research of rhetorical expression, PQP adopts the annotation system with both segmental and suprasegmental aspects for each recording (i.e., each question). Segmental features include three tiers: 1) CHARACTER in Chinese; 2) INITIAL-FINAL (I-F), i.e., the consonant and rhyme (vowel and coda) of the syllable (in Pinyin while distinguishing phonemes); 3) SYLLABLE in Pinyin. For these three tiers, speech data will be first automatically annotated with a custom-written forced-aligned tool, and will be further manually adjusted in Praat.

Suprasegmental features include three tiers that mark at the syntactic word level (except that the construction "demonstrative+classifier" is treated as one word): 4) DESTRESS, i.e., default stress; 5) STRESS, i.e., perceived stress; 6) BOUNDARY, i.e., marking the sentence-final particle. Specifically, Default Stress defines the word that carries the default stress according to the default stress rule [24, 25], which is the last content word in our polar questions. Perceived Stress marks the prosodic prominence level in perception for each word. A native Mandarin speaker with linguistic training, who was blinded to the context of each sentence, was instructed to listen to all sentences and give labels. A three-level stress annotation approach was adopted, represented by 0, 1, and 2, from weak to strong stress. The most prominent word(s) was labeled as 2, and the non-prominent

word(s) as 0. The second most prominent word(s) was labeled as 1 but such a level was not a necessity for each sentence.

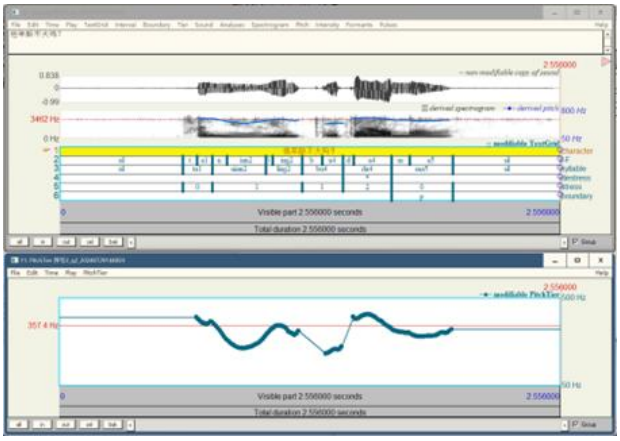

Fig. 1 Upper panel: the audio file and TextGrid file of speech data. Lower panel: the manually corrected PitchTier file.

B. *Spontaneous Responses (SR)*

Via the transcription service provided by the data collection platform Wenjuanxing, each recording of the spontaneous response was converted into a transcription with text and punctuation, which mainly reflects the intonational phrasal boundary of speech. We then performed manual verification of the transcription and identified an utterance as the content between two punctuation marks.

For the audio data, we annotated the segmental features with three tiers for each recording, identical to the system used in PQP. Furthermore, two professionally trained linguistic researchers annotated each response according to Response Intention Annotation System at the utterance level. The third researcher would give a judgement when the former two labeled differently, and all the labels were thoroughly checked by the fourth researcher. Therefore, each response was given a string of intent label(s) that appear in the presence order of utterance. When the two consecutive utterances were considered conveying identical intent(s), one was marked down. Table 2 and 3 illustrate the label definitions and examples.

IV. Discussion and Conclusion

PQSR offers an abundant resource of spoken language (over 8000 recordings from 209 speakers) in the context of question-responding. Moreover, our methodology is both beneficial and generalizable. The experimental design facilitates the documentation of previously less-studied types of conversations, enabling the intentions within these interactions

Tbl. 2 Response Intention Annotation System

Intent Label	Definition
Literal/Factual	***Respond by interpreting the question as inquiring about the confirmation of certain information***
F	a direct answer of yes, no, unsure or "I don't know"
A	a direct answer apparently against the provided contextual information; a lie
T	adding additional conditions, e.g. time, place
W	providing relevant domain knowledge or world knowledge that allows the questioner to derive the answer on their own, or a suggestion or command that enables the questioner to obtain the answer
Attitude	***Respond by interpreting the question as making assertions and expressing dissatisfaction***
S	conveying agreement and submission to the questioner (e.g., apologies)
D	conveying disagreement with the questioner; either by conveying that the questioner should not be dissatisfied given the situation, or communicating that the degree of dissatisfaction should not be as intense
R	to situations where the questioner expresses dissatisfaction towards a third-party; either conveying complete agreement with the questioner, or acknowledging that the questioner has a right to be dissatisfied while simultaneously pointing out that there are reasons for the third-party's actions or behavior
Emotion	***Respond by interpreting the question as releasing (negative) emotion such as worries, anxiety and anger***
E	emotional support, such as comfort, encouragement, or lightening the atmosphere
V	a perfunctory response
Commitment	***Respond by interpreting the question as a command or demand***
I	making a promise
H	demonstrating anticipation or desire
Extension	***Respond to extend the conversation***
C	an indirect response that shifts the topic, which includes changing the focus of attention or diverting the conversation to someone other than the questioner
K	continuation of the original topic raised in the question
M	a counter-question to interrogate the questioner's motive

Tbl. 3 Examples of intention annotation for polar questions (Q-Label) and responses (R-Label)

Polar Question	Q-Label	Response	R-Label
Seeing Andy doing his homework, his mother asks him concernedly: Have you drunk your water? 水喝了吗？	ISQ	I have. Don't worry, mom. 我喝了，妈妈，不用担心。	F E
		I'll drink after I finish this problem. 等我做完这题就去喝。	I
Knowing Andy hasn't drunk water, his mother asks him angrily: Have you drunk your water? 水喝了吗？	RQ	Sorry, mom. I haven't. 不好意思，妈妈，没有喝。	S F
		I don't want to drink water. 我不想喝水。	D
		Mom, I am doing homework. 妈妈，我在写作业呢。	C
Planning to buy a new mobile phone, Emily asks her boyfriend for advice: Does this phone work well? 这个手机好用吗？	ISQ	I'm not really sure, but the reviews online aren't good. 我也不是很确定，但是听网上的评价不太好。	F W
		It depends on what you need it for. 看你要用哪方面的需求了。	T
Knowing Andy hadn't revised and failed the exam, his class teacher asked the subject teacher sarcastically: Did he revise for the exam? 他复习了吗？	RQ	Look at how 'well' he's revised! 看他复习的是真好呀。	A-R
Knowing Andy always skips cleaning duty, his classmate complains to the class monitor: Has Andy done the cleaning duty? 小明做过值日吗?	RQ	No, he hasn't. That's completely unacceptable. 没有呢，太过分了。	F R
Konwing Andy can't dance but still wants to sign up for the dancing competition, his colleague asks the organizer： Can he dance? 他会跳舞吗？	RQ	The experience is what counts。 重在参与嘛。	V
		He said he was pretty good at it when he signed up. Have you ever seen him dance? 他报名的时候说他还蛮擅长的，你见过他跳舞吗？	W K
Hearing the student union is about to inspect the classroom, the class monitor asks the student on duty： Have you cleaned the classroom? 班级的卫生收拾了吗？	ISQ	It's all cleaned up. We'll definitely pass the inspection! 收拾好了，一定能通过卫生检查。	F H
Andy fell asleep in class and he asks his deskmate upon waking up： Is class over yet？现在下课了吗？	RQ	Did you just wake up? 你是不是才睡醒呀？	M

to be traced and analyzed. The naturalistic settings allow for the capture of natural and spontaneous human oral expressions, enriching the content and modes of expression with a highdegree of diversity. In subsequent steps, we will expand the sample size of the corpus to ensure a more even distribution of participants in terms of age, gender, dialect background, and other relevant factors.

By specifically looking into rhetorical questions and their responses, we proposed a framework of communicative intention that emphasizes the respondent's multidimensional considerations concerning the questioner. It thus extends existing theories of speech acts [7-9] by highlighting to what aspects of the questioner the respondent reacts, and distinguishes itself from the existing dialogue corpora [10-13] for applying the brand-new intention framework. We believe that this framework can be generalized to various forms of reactions in human communication, and the annotation system can be customized to specific scenarios through label adjustments. It is noted that direct answers to rhetorical questions with modal verbs (e.g., "Can he dance?") or those related to subjective judgment (e.g., "Is he qualified?") may inherently fall into the dimension of attitude due to the nature of subjective domain. Thus, the impact of the semantics of questions within this framework should be addressed in future research.

This corpus benefits the research of the relationship between intention and the use of prosody. For responses, we rarely observed cases where intention decoded from prosody contradicts intention decoded from the text transcription combined with the contextual information, yet awaiting to be verified. A preliminary study of ISQs and RQs using a subset of this corpus has uncovered the focus shift driven by rhetorical expressions, leading to a global change of intonational contour [26]. In addition, this corpus enables the exploration of how intention conveyance relates to specific linguistic usages in oral conversations, such as using vocatives to address the questioner (e.g., "mum" in "no worry, mum"), discourse markers and various types of gap fillers. Ultimately, we are confident that the PQSR corpus contributes significantly to the advancement of human-computer interfaces for spoken language.

ACKNOWLEDGMENT

We thank Prof. XIONG Ziyu for providing the automatic segmentation and annotation tool and LI Yuanxi for contributing to annotation.

REFERENCES

[1] X. Wang, C. Li, J. Zhao, and D. Yu, "NaturalConv: A Chinese Dialogue Dataset Towards Multi-turn Topic-driven Conversation," Proc. AAAI Conf. Artif. Intell., vol. 35, no. 16, pp. 14006–14014, May 2021.

[2] M. Chen et al., "The JDDC Corpus: A Large-Scale Multi-Turn Chinese Dialogue Dataset for E-commerce Customer Service," in In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, pp. 459–466, March 2020.

[3] K. Zhou, A. Li, Z. Yin, and C. Zong, "CASIA-CASSIL: a Chinese Telephone Conversation Corpus in Real Scenarios with Multi-leveled Annotation," in Proceedings of the International Conference on Language Resources and Evaluation 2010, May 2010, Valletta, Malta.

[4] S. Reddy, D. Chen, and C. D. Manning, "CoQA: A Conversational Question Answering Challenge," in NACCL 2019.

[5] J. Bjerva, N. Bhutani, B. Golshan, W.-C. Tan, and I. Augenstein, "SubjQA: A Dataset for Subjectivity and Review Comprehension," In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, Oct. 2020, pp. 5480–5494.

[6] C.-H. Lee, S.-M. Wang, H.-C. Chang, and H.-Y. Lee, "ODSQA: Open-domain Spoken Question Answering Dataset," in Proceedings of 2018 IEEE Spoken Language Technology Workshop (SLT), 2018, pp. 949-956.

[7] N. Asher and A. Lascarides, Logics of Conversation (Studies in Natural Language Processing). Cambridge University Press, 2005. Accessed: Feb. 17, 2025.

[8] J. R. Searle, Speech Acts: An Essay in the Philosophy of Language, 1st ed. Cambridge University Press, 1969.

[9] K. Samuel, S. Carberry, and K. Vijay-Shanker, "Dialogue Act Tagging with Transformation-Based Learning," In 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Montreal, vol. 2, 1998, pp. 1150–1156.

[10] Y. Zhou, C. Li, B. Xu, J. Xu, L. Yang, and B. Xu, "Constructing a Chinese Conversation Corpus for Sentiment Analysis," in Natural Language Processing and Chinese Computing, vol. 10619, X. Huang, J. Jiang, D. Zhao, Y. Feng, and Y. Hong, Eds., in Lecture Notes in Computer Science, vol. 10619., Cham: Springer International Publishing, 2018, pp. 579–590.

[11] H. Zhang, H. Xu, X. Wang, Q. Zhou, S. Zhao, and J. Teng, "MIntRec: A New Dataset for Multimodal Intent Recognition," in Proceedings of the 30th ACM International Conference on Multimedia, Oct. 2022, pp. 1688–1697.

[12] Z. Zhou, I. G. Choi, Y. He, V. Yadav, and C.-H. Lee, "Using Paralinguistic Information to Disambiguate User Intentions for Distinguishing Phrase Structure and Sarcasm in Spoken Dialog Systems," in 2021 IEEE Spoken Language Technology Workshop (SLT), Jan. 2021, pp. 1020–1027.

[13] Z. Li, G. Zhang, S. Okada, L. Wang, B. Zhao, and J. Dang, "MBCFNet: A Multimodal Brain–Computer Fusion Network for human intention recognition," Knowl.-Based Syst., vol. 296, p. 111826, Jul. 2024.

[14] J. Groenendijk and M. J. B. Stokhof, "Studies on the Semantics of Questions and the Pragmatics of Answers," Dissertation, University of Amsterdam, 1984.

[15] H. Rohde, "Rhetorical questions as redundant interrogatives," May 2006, Accessed: Feb. 17, 2025.

[16] N. Dehé and B. Braun, "The prosody of rhetorical questions in English," Engl. Lang. Linguist., vol. 24, no. 4, pp. 607–635, Dec. 2020.

[17] I. Miura and N. Hara, "Production and perception of rhetoricalquestions in Osaka Japanese," J. Phon., vol. 23, no. 3, pp. 291–303, Jan. 1995.

[18] R. Y.-H. Lo, A. Kiss, and M. Tulling, "The Prosodic Properties of the Cantonese Sentence-Final Particles aa1 and aa3 in Rhetorical Wh-Questions," in Proceedings of the International Congress of Phonetic Sciences, Melbourne, Australia, Aug. 2019, pp. 502–506.

[19] K. Zahner-Ritter, Y. Chen, N. Dehé, and B. Braun, "The prosodic marking of rhetorical questions in Standard Chinese," J. Phon., vol. 95, p. 101190, Nov. 2022.

[20] R. Lo and A. Kiss, "Durational and Pitch Marking of Rhetorical Wh-questions in Mandarin," in Proceedings of the 10th International Conference on Speech Prosody, Tokyo, Japan, May 2020, p. 433.

[21] X. Zhu, "An Analysis of Two Types of Responses," Journal of Xuzhou Normal University, no. 02, pp. 121–124, 1996.

[22] Y. Jun, "Forms and Functions of Verbal Irony in Mandarin Chinese and American English: A Pragmatic Case Study Based on Data From TV Debates," Doctor, Shanghai International Studies University, Shanghai, 2005.

[23] N. Asher, J. Hunter, M. Morey, F. Benamara, and S. Afantenos, "Discourse Structure and Dialogue Acts in Multiparty Dialogue: the STAC Corpus," in Proceedings of LREC 2016.

[24] G. Cinque, "A Null Theory of Phrase and Compound Stress," Linguist. Inq., vol. 24, no. 2, pp. 239–297, 1993.

[25] L. Xu, "Manifestation of informational focus," Lingua, vol. 114, no. 3, pp. 277–299, Mar. 2004.

[26] S. Chen, Q. Sun, Y. Huang, and Y. Luo, "The Prosodic Characteristics of Standard Chinese Rhetorical Questions in Naturalistic Settings," accepted by INTERSPEECH 2025.