INTO-CASS: A CORPUS FOR THE STUDY OF INTONATION AND PROSODY IN CHINESE DIALECTS AND ETHNIC LANGUAGES

Aijun Li, Ziyu Xiong

Institute of Linguistics, Chinese Academy of Social Sciences
liaj@cass.org.cn; xiongzy@cass.org.cn

ABSTRACT

China is well-known for being home to hundreds of related dialects, collectively known as Chinese. These dialects are divided into ten major groups, each of which also contains multiple subgroups, according to *Language Atlas of China* (2nd edition). There are in addition over 100 ethnic languages spoken by approximately 8% of the population of China. While phonological systems were traditionally described in terms of segments, lexical tones or pitch accents, development of speech communication and speech interactive systems calls for a more in-depth understanding of post-lexical prosodic features such as intonation, especially in light of the vast linguistic diversity in Chinese languages. A project has been launched to study the intonation and prosody of Chinese dialects and ethnic languages from typological and interactive perspectives. This paper first proposes a theoretical framework in the study of intonation and prosody, and then introduces the construction of the Corpus for the Study of Intonation and Prosody in Chinese Dialects and Ethnic Languages (INTO-CASS).

Index Terms— Corpus, intonation study, Chinese dialects, ethnic languages in China

1. INTRODUCTION

Prosody is a phonetic phenomenon that occurs above the segmental level and encompasses suprasegmental features including prominence, intonation, rhythm, etc. [1]. Prosody can be observed on both lexical and post-lexical levels. Lexical prosody is basically about prominence, which is realized through tone, stress, and pitch accents, while post-lexical prosody deals with intonation and rhythm.

Chinese is a language that utilizes tones to distinguish lexical meanings. As a result, the study of intonation centers on the interplay of tone and intonation in terms of its phonetic form and pragmatic meaning. Not all languages spoken in China are tone languages. For example, Tibetan belongs to the Sino-Tibetan language family, and it is a tone language like Chinese. Mongolian and Uyghur, both part of the Altaic language family, are different. Uyghur is a lexical pitch-accent language with the location of accent believed to be unpredictable. Whether Mongolian has word-level stress or pitch accent like Uyghur is still under debate.

Studies of Chinese intonation have been prevailingly concentrated upon Mandarin or the "common speech", which is the official variety of Chinese. Research on the Chinese dialects and monitory languages is still in its infancy, but it has shown great potential in making meaningful contributions. It is also promising to expand the study of intonation to incorporate diverse pragmatic functions under a contextualized paradigm. However, there has not been a generally acknowledged research framework for the investigation of intonation systems in Chinese dialects and ethnic languages. Besides, a large-scale corpus that features intonation and prosody of typologically different languages would need to be designed and developed from the perspectives of prosodic typology and speech interaction.

The present paper will first propose a universal framework for studying intonation systems in Chinese dialects and ethnic languages, and then introduces the construction of the Corpus for the Study of Intonation and Prosody in Chinese Dialects and Ethnic Languages (INTO-CASS), with emphases on the design of text prompts and the progress of data collection.

2. A UNIVERSAL FRAMEWORK FOR STUDYING INTONATION IN CHINESE DIALECTS AND ETHNIC LANGUAGES

Among the studies of Chinese intonation, the interplay of tone and intonation is one of the most intriguing topics. This relation was vividly characterized by Chao in two well-known metaphors: the "Rubber Band" Effect and the "Small Wave and Big Wave" Theory. According to the latter, tone and intonation are related in the form of superimposition – either successive or consecutive – just like small waves sitting on top of big waves [2]. Since Chao's seminal works, there has emerged a large body of succeeding research, which includes both descriptive and experimental studies [3, 4, 5, 6]. Hence, our understandings of the physical properties and linguistic functions of Chinese tone and intonation, especially the interaction between tone and intonation, have been constantly advanced.

Informed by the previous studies, a framework for studying intonation in Chinese dialects and ethnic languages

978-1-6654-0870-7/21/$31.00

is established, as demonstrated in Fig. 1. This framework encompasses four aspects of intonation and prosody: 1) lexical prosody, i.e., patterns in the constituents of intonation, which include tone or lexical stress patterns, sandhi patterns of polysyllabic phrases, and sandhi patterns of neutral-tone words; 2) fundamental intonations, focusing on neutral intonation, focal intonation, and declarative and interrogative intonation; 3) contextual intonation, examining the intonation of sentences with modal particles, sentences performing interactive functions, sentences in different discourse structures, and emotional or attitudinal sentences [7]; and 4) phonetic and phonological descriptions across languages and dialects from the typological point of view.

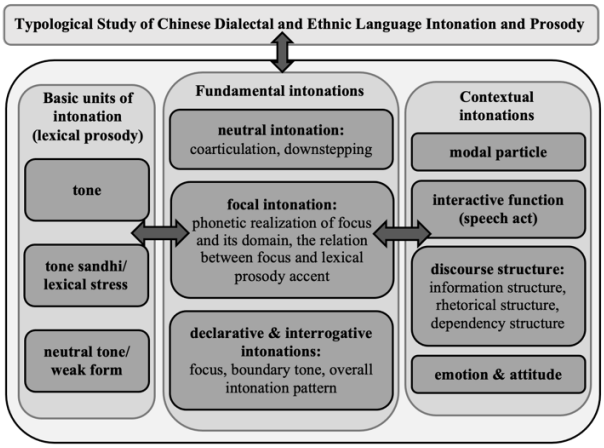

Fig. 1. The universal framework for intonation in Chinese dialects and ethnic languages

3. TEXT PROMPTS

The prompts used in the INTO-CASS construction are divided into Prompts-I for fundamental intonation patterns and Prompts-II for expanded intonation patterns, as listed in Table 1. Prompts-I was compiled in accordance with the framework mentioned above, among which, the prompts of tones, tone sandhi, neutral tones, and basic sentence structures were directly adopted from the materials in dialectal fieldworks [8, 9]. Prompts-II were developed from studies in modern Chinese and ethnic languages. These prompts were designed to investigate the possible connection of form and function of intonation and prosody in Chinese dialects and ethnic languages.

Table 1. Prompt list.

Prompt-I (fundamental intonation patterns)		**Prompt-II (expanded intonation patterns)**	
Basic units of intonation (W)	• Tone (W1) • Tone sandhi (W2-W3) • Neutral tone (NW)	50 core syntactic structures (S2)	
Basic intonation patterns (S1)	Statements and yes-no questions with target words	Interactive functions (S3)	• Colloquial sentences (S3-1) • Emotional sentences (S3-2)
	and focus-controlled (S1-1)		
	Target words embedded in sentences with various syntactic structures (S1-2)	Sentences with structural ambiguity (S4)	
		Narratives (D)	

3.1. Design of Prompts-I

3.1.1. Basic units of intonation (W)

The prompts of basic units of intonation were intended to examine acoustic patterns of basic intonation constituents. Thus, the prompts include the word lists for monosyllabic tones (W1), disyllabic tone sandhi (W2), and trisyllabic tone sandhi (W3). But, for those ethnic languages without tone such as Mongolian or Uyghur, the word list should cover various word stress or word accent or lexical structures. W1~W2 will be described in detail.

3.1.1.1. Word list for monosyllabic tones (W1)

The classification of tone categories follows the four-tone system in Middle Chinese as "*Ping, Shang, Qu, Ru*", which is supposed to be an eight-category system according to the voicing of the initial as "*Yin*" for upper register and "*Yang*" for lower register. The eight categories are then *yin-ping*, *yang-ping*, *yin-shang*, *yang-shang*, *yin-qu*, *yang-qu*, *yin-ru*, and *yang-ru*. Tones of modern Chinese in each dialect are evolved from this eight-category system.

The word list for monosyllabic tones, i.e., the citation tone for each character (W1) is adopted on the basis of the word list for investigating Middle Chinese tones in *The Word List for Chinese Dialectal Fieldwork*. W1 will be used in the recording for each dialect so as to compare tone patterns across dialects, and examine the correspondence between modern tones and their ancient counterparts, shown as Table 2 below.

Table 2. Word list for monosyllabic tones (W1).

tones	**Onset**		**Example characters**
Ping	*Yin*	unaspirated voiceless	东 该 灯 风
		aspirated voiceless	通 开 天 春
	Yang	voiced sonorant	门 龙 牛 油
		voiced obstruent	铜 皮 糖 红
Shang	*Yin*	unaspirated voiceless	懂 古 鬼 九
		aspirated voiceless	统 苦 讨 草
	Yang	voiced sonorant	买 老 五 有
		voiced obstruent	动 罪 近 后前～
Qu	*Yin*	unaspirated voiceless	冻 怪 半 四
		aspirated voiceless	痛 快 寸 去
	Yang	voiced sonorant	卖 路 硬 乱
		voiced obstruent	洞 地 饭 树
Ru	*Yin*	unaspirated voiceless	谷稻～ 百 搭 节 急
		aspirated voiceless	哭 拍 塔 切 刻
	Yang	voiced sonorant	六 麦 叶树～ 月
		voiced obstruent	毒 白 盒 罚

3.1.1.2. Word list for disyllabic tone sandhi (W2)

Tone sandhi is the phonological alteration of tones, resulting from two tones occurring in juxtaposition. The third tone sandhi in Mandarin is the best-known example. In contrast, tonal coarticulation involves more subtle changes, not triggering alternations in tonal categories, due to overlapping of distinct articulatory gestures.

Tone sandhi patterns are very complex in Chinese dialects. It can be categorized according to direction, manifestation, and domain of sandhi. For example, the pattern of the third tone sandhi in Mandarin is right-dominant in that Tone 3 on the right-most syllable triggers the tonal alteration of the preceding syllable. The tone sandhi pattern in Shanghainese and Northern Wu dialect is left-dominant, which means the left-most syllable in a lexical compound maintains its citation tone and spreads its tone over the domain [10]. Furthermore, the tone sandhi pattern in the Southern Min dialect as spoken in Xiamen, also known as Amoy, is tonal substitution. Namely, while the right-most syllable maintains its citation tone, other syllables in the domain are circularly replaced by one another, forming a pattern of alternation, often known as the Min Tone Circle [11].

Therefore, when it comes to the design of the word list for disyllabic tone sandhi (W2), the sandhi domain is defined as the combination of two syllables. W2 covers all possible permutations of eight tone categories in disyllabic combinations. Four entries are selected for each permutation with different morphological structures taken into consideration. Hence, there are 8*8*4=256 combinations of disyllabic words in total. Given space limitation, W2 is omitted in this paper.

3.1.1.3. Word list for trisyllabic tone sandhi (W3)

In the similar vein, there are 8*8*8*4=2048 combinations of trisyllabic words in the word list for trisyllabic tone sandhi (W3), covering four entries of each permutation of eight tone categories in each syllable in the trisyllabic words. Words in both 2+1 and 1+2 structures were considered. W3 is also omitted in this paper due to space limitation.

3.1.2. Basic intonation patterns (S1)

3.1.2.1. Intonations patterns of statements and yes-no questions with different focus structures (S1-1)

Focus structure is an integration of the information structure triggered in a certain discourse context, the syntactic structure of the utterance, and its intonational structure [12]. Focus plays a role of highlighting the message intended by the speaker. In terms of phonetic realization, it is often associated with the sentence-level stress, which is a key topic in the study of intonation. Xu reported the post-focus compression (PFC) in that the pitch range preceding the focused element remain unaltered, but that of the focal syllable(s) is significantly amplified, with the post-focus pitch range greatly compressed [13]. He further found out that the lowering of pitch in PFC is a gradient drop of F0 following the sentence topic [14]. Besides, the phonetic realization of different types of foci can be seen in Jia and Li's work [15]. However, Chinese dialects is different from Mandarin in the phonetic realization of focus. For example, while PFC is found prevalent among Mandarin dialects, Cantonese and Min dialect exhibit a lack of this phenomenon.

When it comes to the design of prompts, the focal sentences are constructed, following the word order of Subject–Predicate–Modifier–Object. Accordingly, the basic sentence adopted is *Zhāng jūn chūsān fēi tiānjīn?* (Jun Zhang is going to fly to Tianjin on the third of the lunar calendar?). The target words are disyllabic words covering all permutations of eight tone categories. They occur in the subject, modifier, and object positions respectively. Three types of focus, i.e., information focus, contrastive focus, and broad focus, are considered. Hence, there are 8*8*3*3=576 entries of declarative sentences and 576 entries of interrogative sentences in total.

Additionally, in order to obtain focus information, all the utterances are presented within a dialogue. The response in (1) is a statement bearing the information focus. (2) presents a question and a statement of the contrastive focus, in which the subject bears the focus. Focus is indicated in bold.

(1) *Shuí chūsān fēi tiānjīn?* ***Zhāngjūn*** *chūsān fēi tiānjīn.*
"Who is going to fly to Tianjin on the third of the lunar calendar? **Jun Zhang** is going to fly to Tianjin on the third day of the month."

(2) ***Sānjiù*** *chūsān fēi tiānjīn?* ***Āyí*** *chūsān fēi tiānjīn.*
"Is **Third Uncle** going to fly to Tianjin on the third day of the month? **Aunt** is going to fly to Tianjin on the third day of the month."

3.1.2.2. Intonation patterns of various common sentence structures (S1-2)

Besides statements and yes-no questions, prompts of more sentence structures were compiled to further examine the interplay of lexical prosody and intonation with diverse syntactic and prosodic structures.

Four target words were selected from everyday vocabulary in dialects, which were *jītāng*, "chicken soup", *yútāng*, "fish soup", *mǐtāng*, "rice soup", and *miàntāng*, "noodle soup". They were inserted into a variety of common sentence patterns that include statements, questions, exclamations, commands, as well as contrastively focused sentences, topic sentences, comparative sentences, and sentences of various rhetorical relations, some of which are shown in (3).

(3) a. *Tā ài hē de shì jītāng, bú shì yútāng.*
"What he likes to drink is chicken soup, but not fish soup."
b. *Tā ài hē de bú shì jītāng, shì yútāng.*
"What he likes to drink is not chicken soup, but fish soup."
c. *Jītāng bǐ yútāng gèng hǎo hē?*
"Chicken soup tastes better than fish soup?"
d. *Jītāng bǐ yútāng gèng hǎo hē ma?*
"Does chicken soup taste better than fish soup?"

A total of 95 different common syntactic structures were created, with four target words inserted in each of them. Hence, there are 95*4=380 entries in this part.

3.2. Design of Prompts-II

3.2.1. Intonation patterns of core syntactic structures (S2)
50 sentences were selected from the grammar investigation part in *Project for Protection of Language Resources of China* to examine the intonation patterns of the core syntactic structures in Chinese dialects and ethnic languages [8].

3.2.2. Intonation patterns for interactions with various syntactic structures (S3)
137 entries of colloquial sentences or dialogues (S3-1) were selected from the classic works of Chinese linguists [16, 17]. To illustrate, (4) examines two usages of the aspect particle *le* ("le_1" is used after the verb to indicate the completion of the action. "le_2" is used at the end of a sentence to affirm that the situation has changed or is about to change.), and the usage of the negative form of the modal *bú yòng*.

(4) *Tā zǎo chī le_1 fàn le_2, bú yòng děng tā.*
"He has had a meal already, (so) there is no need to wait for him."

Sentences in (5-9) examine the usage of a variety of modal particles.

(5) *Dōu xiǎodiǎn shēng a, háizi shuì le.*
"All be quiet. The kid has been to bed."
(6) *Wǒ shízài shì gēn tā guò bú xiàqù le a!*
"There's really no way for me to get along with him!"
(7) *Nǐ xiān zǒu ba, wǒ suíhòu jiù lái.*
"You can go first, and then I'll come."
(8) *Nǐ zěnme la? Bìng la?*
"What's going on with you? Are you sick?"
(9) *Guò nián la, tài kāixīn le.*
"It's the Lunar New Year. I'm so happy."

The dialogues in (10) - (12) examine the interrogative function of *wh*-words and their non-interrogative function of arbitrary reference.

(10) A: *Nǐ hē shénme?*
"What would you like to drink?"
B: *Shénme dōu xíng, suíbiàn ba.*
"Anything will do, whatever."
(11) A: *Nǐ zhǎo shuí ne?*
"Who are you looking for?"
B: *Lǎo Zhāng.*
"Zhang."
(12) A: *Shì shuí zuótiān lái de?*
"Who came yesterday?"
B: *Shì lǎo Zhāng.*
"It was Zhang."

There are also 50 entries of emotional sentences (S3-2) in this part, in which emotional words are present as in (13) and (14), or absent as in (15) and (16).

(13) *Wǒ shízài tài kāixīn le!*
"I'm really extremely happy!"
(14) Qì sǐ wǒ le!
"I'm pissed off!"
(15) *Tāmen qíchē dào túshūchéng.*
"They ride a bicycle to the Book City."
(16) *Yúxiāngròusī.*
"Shredded pork in spicy sauce."

3.2.3. Intonation patterns of sentences with structural ambiguity (S4)
20 sentences with structural ambiguity that can be disambiguated by differences in prosodic structure are laid out as S4. Examples are given in (17) and (18), in which phrasing is indicated by "/".

(17) a. *Xīn/ jiào shī xù shè/ yǐ jīng luò chéng.*
"New faculty dormitory has been completed."
b. *Xīn jiào shī/ xù shè/ yǐ jīng luò chéng.*
"The dormitory for new faculty has been completed."
(18) a. *Pīng pāng qiú/ pāi mài wán le.*
"The auction of table tennis rackets is over."
b. *Pīng pāng/ qiú pāi/ mài wán le.*
"The table tennis rackets are sold out."

3.2.4. Narrative speech (D)
The text prompts include *The North Wind and the Sun*, *The Cowherd and the Weaving Maid*, and *The Pear Story*. In this part, the speakers will be asked to firstly read and then recount these stories in natural speech.

3.3. Adapting prompts for specific dialects or ethnic languages

Prompts-I and Prompts-II are originally a universal design scheme. It is necessary to adapt the universal prompts to the phonological, morphological, and grammatical systems of specific dialects or ethnic languages before recording. For example, once the tone system in a specific dialect has been affirmed, the number of the tone sandhi combinations will be adjusted accordingly. Thus, the disyllabic and trisyllabic word lists can be simplified to include the combinations existing in this dialect. Shanghainese, for example, contains five tones, and the combinations of disyllabic and trisyllabic word lists can be reduced to 25 and 125 respectively. In addition, the vocabulary and sentences in universal prompts may be different in each dialect or ethnic language. Thus, the prompts need to be adapted through paraphrasing or translating these words or sentences into corresponding dialects or ethnic languages, as illustrated in (19).

(19) a. General prompts:
Tā xǐhuān hē de shì jītāng, bú shì yútāng.
"What he likes to drink is chicken soup, but not fish soup."
b. Shanghainese prompts: *Yī huānxǐ chī de shì jītāng, wù shì yútāng.*
"What he likes to drink is chicken soup, but not fish soup."

3.4. Supplementary prompts for special prosodic phenomena

Special prosodic phenomena in some dialects and ethnic languages should be treated as supplementary prompts for further studies. For instance, Chengdu dialect is rich in reduplicative words in neutral tone. Accordingly, reduplicative words in different parts of speech are chosen as target words and inserted into focal sentences, as in (20).

(20) *Māma tiāntiān zuò héhe?*
"Mother makes boxes every day?"

Similarly, Cantonese exhibits a rich system of modal particles, including approximately 40 frequently used sentence-final modal particles, some of which are disyllabic. Unlike Mandarin, Cantonese modal particles are typically said in one of the lexical tones rather than in neutral tone. Hence, we designed sentences such as (21).

(21) A：*Dim2gaai2 nei5 gam1yat6 gam3 mong4 ge3.*
"Why are you so busy today?"
B：*Gu1ma1 gam1jiu1 faan1 Sou1jau1 lo1.*
"Aunt is going back to Suzhou this morning."

4. RECORDED DATA

4.1. Criteria for selection of dialects and ethnic regions

The goal of INTO-CASS is to collect the speech data from ten major dialects and ethnic languages in China, with close attention paid to the complexity of lexical prosody (tone categories, tone sandhi, word stress, or word accent). For example, Honggu dialect of Northwestern Mandarin has merely two tones, whereas Guangzhou dialect of Cantonese has nine tones, and Changli dialect of Jilu Mandarin has no level tone, but only contour tones [18]. Moreover, the sandhi pattern of Taifeng dialect, a subdialect of Wu, is conditioned by morphology and the right-dominant pattern whereas Shanghainese is left-dominant, and the tones in Changzhou dialect are distributed in either high or low pitch register [19]. In the current phase, nine dialectal regions and three ethnic regions (Mongolian, Tibetan, and Uighur) have been selected for the purpose of data collection. See Appendix I for the regions selected and recording prompts for each region.

4.2. Speakers and recoding

The speakers are split into two groups, middle-aged and elderly (aged above 45) and youth (aged below 45), at least 10 for each group will be recorded. They are expected to use the dialect or ethnic language in daily communication and have no hearing or speaking impairment. The data were recorded and checked by the assistants who can speak or understand the dialect or ethnic language recorded. After checking, the speakers were asked to record the unqualified sentences again. As the recording and corpus construction are still on-going and planned to be completed by the end of 2021, Appendix II only shows the work that has been finished.

4.3. Annotation and studies on INTO-CASS

All speech data are annotated segmentally and prosaically [20], as shown in Fig. 2, where six layers are annotated: sentence in Chinse characters, words in Chinese characters, syllables, initials and finals, prosodic boundary, and stress levels.

A considerable amount of studies based on INTO-CASS have been carried out, including those on tone and tone sandhi in Taifeng Wu dialect [19, 21], Changli dialect of Jilu Mandarin [22, 23], and Wuzhi Jin dialect; those on intonations of three dialects: Cantonese [18,24,25], Honggu [26], and Changzhou; and that on the word stress of Mongolian.

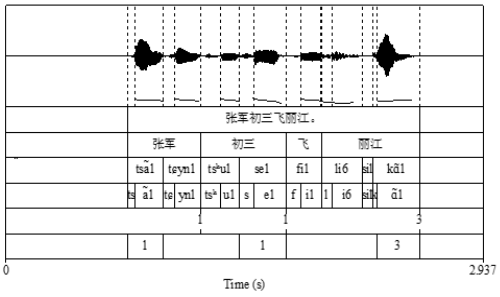

Fig. 2. An Annotation Example for shanghai dialect.

5. SUMMARY

In conclusion, this paper outlines the research framework and construction practice of INTO-CASS. Informed by previous studies of intonation and prosody in Mandarin, the universal framework examines intonation in Chinese dialects and ethnic languages, from constituent units to variation patterns, along with exploration of the interplay of tone and intonation as well as consideration of its interactive and discourse functions. The construction of INTO-CASS is described in terms of the design of universal prompts, the adaption of prompts to specific dialects and ethnic languages, and the current recording in process. INTO-CASS is the first large-scale corpus of intonation and prosody in China. More work on the study of intonation in Chinese dialects and ethnic languages and the collection of prosody-rich data are expected to happen in the near future.

6. ACKNOWLEDGEMENTS

This work is supported by the "Four Batches" Talent Project "Intonation typology". The contribution provided by the members of the project, Zhiqiang Li, Aomin, Mengxue Cao, Hongliu Jiang, Haibo Wang, Yuzhu Zhu, Xiaoyan Zhang, is greatly acknowledged.

7. REFERENCES

[1] H. Fujisaki, "Prosody, Models, and Spontaneous Speech," In Sagisaka, Y., N. Campbell, and N. Higuchi (Eds.), *Computing Prosody*, Springer, New York, pp. 27-42, 1997.
[2] Y-R. Chao, "Tone and Intonation of Chinese," *Bulletin of the Institute of History and Philology, 4*(3), pp. 121-134, 1933.
[3] Wu, Z., *Wu Zongji Yuyanxue Lunwenji [Zongji Wu's Collected Papers on Linguistics]*, The Commercial Press, Beijing, 2004.
[4] J. Shen, "Hanyu Yudiao Moxing Chuyi [An Attempting Discussion of Chinese Intonation Model]," *Yuwen Yanjiu [Linguistic Research], 4*, pp. 16-24, 1992.
[5] Y. Xu, "The PENTA Model of Speech Melody: Transmitting Multiple Communicative Functions in Parallel," *Proceedings of from Sound to Sense, 50*, pp. 91-96, 2004.
[6] Lin, M., *Hanyu Yudiao Shiyan Yanjiu [Experimental Study on Chinese Intonation]*, China Social Sciences Press, Beijing, 2012.

[7] K. Zhou, A. Li, and C. Zong, "Dialogue-Act Analysis with a Conversational Telephone Speech Corpus Recorded in Real Scenarios," *Proceedings of OCOCOSDA 2010*, Katmandu, Nepal, 2010.
[8] Department of Language Information Management of Ministry of Education of the PRC, and Center for the Protection and Research of Language Resources of China, *Zhongguo Yuyan Ziyuan Diaocha Shouce· Hanyu fangyan [The Handbook of Investigating Chinese Language Resources· Chinese Dialects]*, The Commercial Press, Beijing, 2015.
[9] Institute of Linguistics, Chinese Academy of Social Sciences, *Fangyan Diaocha Zibiao [The Word List for Chinese Dialectal Fieldwork]*, Science Press, Beijing, 1981.
[10] Y. Chen, "Revisiting the Phonetics and Phonology of Tone Sandhi in Shanghai Chinese," *Speech Prosody 2008*, Campinas, Brazil, pp. 253-256, 2008.
[11] M. Y. Chen, "The Syntax of Xiamen Tone Sandhi," *Phonology*, *4*, pp. 109-149, 1987.
[12] Lambrecht, K., *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press, Cambridge, 1996.
[13] Y. Xu, "Effects of Tone and Focus on the Formation and Alignment of F0 Contours," *Journal of Phonetics, 27*(1), pp. 55-105, 1999.
[14] B. Wang, Y. Xu, and J. Xu, "Prosodic Realization of Discourse Topic in Mandarin Chinese: Comparing Professional with Non-professional Speakers," *Proceedings of 17th ICPhS*, Hong Kong, China, pp. 2086-2089, 2011.
[15] Y. Jia, and A. Li, "Phonetic Realization of Accent from Chinese English Learners in Various Dialectal Regions," *ISCSLP2012*, IEEE, Hong Kong, China, pp. 296-300, 2012.
[16] Lv, S., *Xiandai Hanyu Babai Ci [Eight Hundred Words in Modern Chinese]*, The Commercial Press, Beijing, 1980.
[17] Y-R. Chao, "Beiping Yudiao de Yanjiu [A Study on Peking Intonation]," In Wu, Z., and X. Zhao (Eds.), *Zhao Yuanren Yuyanxue Lunwenji [Yuen-Ren Chao's Collected Papers on Linguistics]*, The Commercial Press, Beijing, pp. 253-272, 2002.
[18] Ge, C., Hanyu Yudiao de Kuafangyan Yanjiu [A Cross-Dialectal Study on Chinese intonation] [Doctoral thesis, University of Chinese Academy of Social Sciences], 2020.
[19] X. Zhang, A. Li and Z. Li, "Complex Patterns of Tonal Realization in Taifeng Chinese," *ISCSLP 2021*, Hong Kong, 2021.
[20] A. Li, "Chinese Prosody and Prosodic Labeling of Spontaneous Speech," *Speech Prosody 2002,* Aix-en-Provence, France, pp. 39-46, 2002.
[21] A. Li, X. Zhang, and Z. Li, "Acoustic and Phonological Analyses of Tones in Taifeng Chinese," 10th *International Conference on Speech Prosody*, Tokyo, Japan, pp.655-649, 2020.
[22] M. Cao, "Changli Chengguan Fangyan Yuyin Tanwei [Exploring the Phonetics of Changli Chengguan Dialect]," *Chinese Journal of Phonetics, 13*, pp. 81-89, 2020.
[23] M. Cao, "Hebei Changli Chengguan Fangyan Feidiezi Liangzizu Liandu Biandiao he Qingsheng [The Disyllabic Tone Sandhi of Non-reduplication and Neutral Tones of Hebei Changli Chengguan Dialect]," *Fangyan [Dialect], 4*, pp. 401-410, 2020.
[24] C. Ge, and A. Li, "Declination and Boundary Effect in Cantonese Declarative Sentence," *ISCSLP 2018*, Taipei, 2018.
[25] C. Ge, and A. Li, "Intonation of Cantonese Interrogative Sentences with and without Sentence Final Particle," *ICPhS 2019*, Melbourne, Australia, pp. 2455-2459, 2019.
[26] C. Ge, and A. Li, "Tonal Variations in Honggu Chinese," *Speech Prosody 2020*, Tokyo, Japan, pp. 245-249, 2020.

Appendix I. An overview of prompts for each region (M. for Mandarin, SW for Southwest, & NW for Northwest).

Regions	**Standard M.**	**SW M.**		**Jilu M.**	**NW M.**	**Wu**			**Cantonese**	**Jin**	Mongolian
Cities	Beijing	Chengdu	Qionglai	Changli	Honggu	Shanghai	Changzhou	Taifeng	Guangzhou	Wuzhi	Baotou
W1	348	/		/	40	64	35	81	/	64	/
W2	1251	100	100	/	/	128	/	308	/	/	379
W3	2048	128	128	138	98	/	98	/	/	894	/
NT	50	/	/	/	/	/	/	/	/	/	/
S1-1	44	204	204	112	/	168	/	570	/	74	54
S1-2	131	234	234	124	90	183	180	570	108	74	54
S2	50	190+140	190	162+132	/	50	/	/	/	380	108+18
S3-1	143	156	/	163	/	190+136	/	/	85	200	324
S3-2	/	163	/	163	/	137	/	/	/	50	/
S4	63	49	/	49	/	109	/	/	/	20	/
D	/	2	2	2	/	58	/	/	/	1	/
Total	*4128*	*1366*	*858*	*1045*	*228*	*1223*	*313*	*1529*	*193*	*1757*	*937*

Appendix II. Number of speakers recorded.

Speakers	Gender	Beijing	Chengdu	Qionglai	Changli	Honggu	Shanghai	Changzhou	Taifeng	Guangzhou	Wuzhi	Baotou
Middle-aged & elderly	Male		1		5	3	5	4	10		4	
	Female		3		5	3	5	4	7		2	
Youth	Male	1	3		5		6	1	5	7	2	1
	Female	1	5	1	5		6	1	3	7	2	2
Total		*2*	*12*	*1*	*20*	*6*	*22*	*10*	*25*	*14*	*10*	*3*