Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Cross-linguistic Influence on Intonation Acquisition: A Study on the Production of L2 Mandarin and L3 English Intonations by Uyghur Speakers

Tong Li, Hui Feng

Tianjin University, China
litong_lt@tju.edu.cn, fenghui@tju.edu.cn

Abstract

Intonations in different languages serve the universal function of conveying communicative information and expressing affective meaning, while the prosodic encoding of the same intonational meaning is language-specific. For native speakers of Uyghur in China learning Mandarin as L2 and English as a third language (L3), few studies have been on their cross-linguistic acquisition of prosodic encoding of interrogative and declarative intonations in non-native languages.

Applying the L2 Intonation Learning theory (LILt), this study investigated the prosodic patterns of declarative and interrogative intonations in Uyghur, Mandarin, and English by 20 Uyghur speakers and compared them with 20 L1-Mandarin speakers and 10 L1-English speakers. Findings showed that the prosodic encoding of Mandarin intonation by Uyghur speakers was assimilated to Uyghur. When the pitch competition between lexical tones and interrogative intonation occurred, Uyghur speakers prioritized the prosodic encoding of intonation rather than the lexical tones, resulting in the gradual rising pitch level and rising boundary tones in Mandarin questions with the dipping tones (T3) and falling tones (T4). Regarding their English intonation, the delayed divergence between the downward declarative and upward interrogative intonations suggests that Uyghur speakers relied more on the sentence-final prosody in L1-like norms to distinguish the two intonations.

Index Terms: intonation acquisition, L1-Uyghur speakers, cross-linguistic influence

1. Introduction

In many of the world's languages, statements and questions are distinguished by the contrast of falling and rising intonations [1-4]. However, the phonetic realization and phonological representation vary across languages, either locally or globally within the utterance [3, 5]. The learning of prosodic encoding of intonational meaning not only reduces the non-native accentedness but also improves the efficiency of cross-linguistic communication. Previous studies on intonation acquisition focus on bilinguals; nevertheless, few investigate both L2 and L3 intonation production by trilinguals.

The present study examines Mandarin Chinese (hereinafter Mandarin) and English intonations produced by Uyghur speakers in China who learn Mandarin as L2 and English as a third language (L3). As a member of the Turkic branch of the Altaic language family [6], Uyghur is a non-tone language with regular stress located at the word-final syllable [7-9]. In contrast to Uyghur and English, Mandarin is a typical tone language with four lexical tones and a neutral tone. The three languages also differ in the prosodic contrast between declarative and interrogative intonations. In Uyghur, the rising pitch of the sentence-final syllable plays a dominant role in expressing the interrogative meaning [10, 11]. In English, different temporal scopes of the acoustic contrast between the two intonations have been proposed: at sentence final [12] or over the larger domain [13, 14]. As to Mandarin, the sentence-final syllable adheres to the original tone shape with higher pitch register and shorter duration in the interrogative [15, 16]. Given the diversity of prosodic systems and intonational prosodic encoding across the three languages, it would be interesting to investigate the cross-linguistic influence on the production of both Mandarin and English intonations by Uyghur speakers.

Previous studies pointed out that Uyghurs are highly sensitive to perceiving the final rise of Mandarin interrogative intonation [17] and face more difficulty manipulating the tonal pitch range [18]. Little attempts have been made to specify how the non-tone native language facilitates or affects the production of L2 Mandarin and L3 English intonation and whether L2 influence overwhelms L1 on L3 intonation.

To specify the source and direction of transfer in L2 and L3, this study adopts the framework for cross-linguistic comparison of intonation provided by L2 Intonation Learning theory (LILt) [19, 20]. It includes (1) *systemic dimension*: the inventory of the phonological categories of pitch events such as pitch accent, boundary tones, and the patterns of their combination; (2) *frequency dimension*: the frequency of the occurrence of phonological intonation pattern; (3) *realizational or phonetic dimension*: the phonetic implementation of phonological elements including F0 range, F0 alignment, mean F0, etc. and (4) *semantic dimension*: how the distribution of elements is associated with intonational meaning. In this particular study, we will take the systemic and frequency dimensions together as the phonological dimension and focus on the boundary tones. The realizational and semantic dimensions are explored, concentrating on the differentiation of prosodic encoding patterns between declarative and interrogative intonations.

LILt also proposes that the influence from L1 takes the form of assimilation or polarization in L2 production, and the cross-linguistic influence may occur in any dimension [19, 20]. We assume that it is also applicable in L3 intonation. Therefore, this research aims to address the following questions: (1) How does the production of declarative and interrogative intonations in Mandarin and English by Uyghur speakers differ from native speakers? (2) How does the cross-linguistic influence contribute to the acquisition of Mandarin and English intonations respectively?

2. Methodology

2.1. Participants

A total of 20 native speakers of Uyghur (NSU), 20 native speakers of Mandarin (NSM), and 10 native speakers of English (NSE) participated in this experiment. Uyghur speakers were

10.21437/SpeechProsody.2024-20

ten male and ten female university students (average age = 22.0; *SD* = 1.6) born and raised in Ürümqi. They have learned Mandarin for 14 years (*SD* = 2.3) and English for four years (*SD* = 1.0). Their Mandarin proficiency is at B2 level and English proficiency achieves A2 according to the CEFR. Mandarin speakers were from northern China including ten male and ten female students (average age = 22.5; *SD* = 1.6), and four male and six female native speakers of English (average age = 30.7; *SD* = 5.6) were born and raised in the USA. Both NSM and NSE speak the respective standard language varieties without self-reported or perceivable accents.

2.2. Data recording and procedures

To elicit intonations that are comparable across languages, three sets of Uyghur, Mandarin, and English sentences are designed, and each set includes four sentences that are composed of four disyllabic words (examples given in Table 1). The reading texts of Mandarin have different tone sequences (T1, T2, T3, T4). Considering the stress-driven distribution of pitch accent in English, words in two sentences all have initial lexical stress and are final-stressed in the other two.

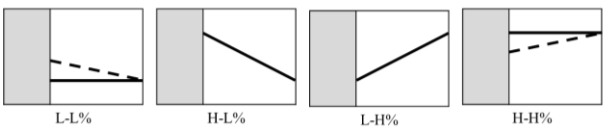
Table 1: *Uyghur, Mandarin and English sentences.*

Language	Sentence examples
Uyghur	naːdɛm aχʃam kino kørdi *(Nadam watched movie last night)*
Mandarin	Zhang1bin1 jin1tian1 can1guan1 gong1si1 *(Zhangbin visited the company today)*
English	Mary cancelled orders lately

The program SpeechRecorder [21] was used to display the target sentence on the screen, one at a time, in random order for each speaker. Each sentence was prompted to two sentence types (statement and yes/no question) by adding a period or a question mark to the end. The speech was recorded in a quiet room and digitized at a 44.1kHz/16bit sampling rate. Ultimately, 160 tokens of Uyghur intonations from NSU (20 participants × 4 sentences × 2 types), 320 tokens of Mandarin intonation from NSU and NSM (40 participants × 4 sentences × 2 types), and 240 tokens of English intonation from NSU and NSE (30 participants × 4 sentences× 2 types) were collected.

2.3. Data processing

Recordings were annotated manually in Praat [22]. For better cross-linguistic comparison of prosodic features, only sentence final boundary tones were annotated with the same standards. A simplified representation for each type of boundary tone is illustrated in Figure 1.



Note. The grey parts represent the accented syllables and the white parts represent unaccented syllables. The dotted lines represent the alternative pitch curves.

Figure 1: *Schematic representation of boundary tones.*

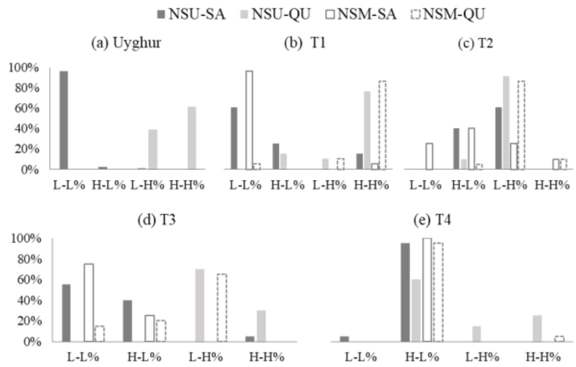
For phonological analysis in systemic and frequency dimensions, the types of boundary tones appearing in Uyghur, Mandarin, and English by different speakers were counted, and the corresponding frequencies were calculated. For phonetic analysis in the realizational dimension, F0 values in Hertz (Hz) in each syllable were extracted at ten points in equal intervals and were converted to semitones (st) where the reference frequency was set as 1 Hz. In addition to the direct comparison of pitch movement, linear discriminant analysis (LDA) was conducted in R [23] based on the local and global phonetic features of intonation extractd by Praat scripts to predict the relationship between various phonetic features and intonational meaning by each speaker group [24]. Global features include the mean duration per syllable, mean F0, F0 range and slope of linear regression of the whole sentence, and the local features are the duration, mean F0 and F0 range of each word.

3. Results and discussion

3.1. L1 influence on Mandarin intonation acquisition

3.1.1. Boundary tones

The inventory and frequencies of boundary tones in Uyghur intonations by NSU and Mandarin intonations by NSU and NSM were summarized in Figure 2.



Note. SA= statement; QU = question.

Figure 2: *The inventory and frequencies (%) of IP boundary tones in Uyghur and Mandarin intonations*.

Three-way ANOVAs and post-hoc Scheffe Tests were conducted to compare the frequencies of four boundary tones in Uyghur intonations by NSU, Mandarin intonations by NSM, and Mandarin intonations by NSU. When all syllables bear T1 in the question, no significant difference was found between NSU and NSM. For T2 sequence, the significant high frequency of rising tone in NSU implies the difficulties in reducing L-H% in Mandarin T2 statement ($p < .001$). NSU were proficient in L-H% and H-H% in T2 question since no differentiation was demonstrated compared to NSM. In T3 sequence, when in the declarative intonation, NSU tended to employ high frequency of H-L% as NSM, while the much higher frequency of H-H% in the interrogative ($p < .01$) was similar to Uyghur. In T4 statement, similar to NSM, NSU used H-L% as the majority of boundary tones. However, for the interrogative intonation, affected by Uyghur, L-H% and H-H% were used with significantly higher percentages ($p < .05$; $p < .01$).

Given the absence of specific tonal specifications for each T1 syllable, coupled with the similarity of the rising pitch between T2 and Uyghur questions, NSU were easier to learn the boundary tones in T1 and T2 questions. However, T3 is a contour tone with a low dipping and T4 is featured by high-falling tonal shape. In T3 and T4 questions, NSM preserved the L tonal target and falling tone respectively. In contrast, the motivation of NSU to raise the final intonation overwhelmed the preservation of T3 and T4 tonal shapes, which were influenced by L1 where final rising was a significant indicator of interrogative meaning. Therefore, for T3 and T4 questions, L1 influence led to the assimilation of the prosodic encoding of Mandarin intonations by NSU to the L1 intonation.

3.1.2. Prosodic encoding patterns

From Figure 3 and Figure 4, the pitch contours of Uyghur and Mandarin intonations by NSU were compared to NSM with different tone conditions (SYN refer to the Nth syllable from the left in sentence). Uyghur declarative and interrogative intonations began at nearly identical pitch levels and the most distinctive pitch difference lay in the sentence-final syllable, that is, the marked expansion of the pitch range ($t(19) = 5.076$, $p < .001$) and increase of the pitch level ($t(19) = 8.886$, $p < .001$), which was consistent with the findings of [10].

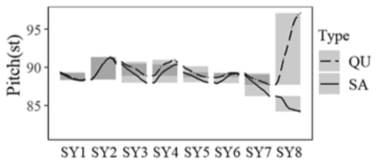


Figure 3: *Pitch curves of Uyghur intonation by NSU.*

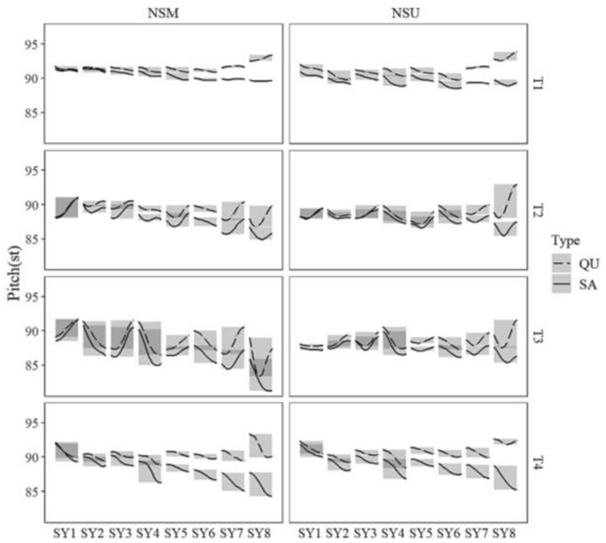


Figure 4: *Pitch curves of Mandarin intonations by NSM and NSU.*

In T1 and T4 sequences, NSU had the same trends of pitch movement as NSM, that is, the pitch level fell in declarative intonation and rose in interrogative intonation. In T2 and T3 sequences, the same pattern was also adopted by NSU. However, NSM had the gradual falling pitch in both declarative and interrogative intonations. Local differences in the final intonation assimilated to Uyghur were also found. For example, the pitch of the last syllable in T2 interrogative intonation increased dramatically ($t(19) = 5.643$, $p < .001$), which was different from the general falling pitch trends by NSM. SY8 in the T4 interrogative did not preserve the original HL tone and had a significantly smaller pitch range ($t(19) = 7.72$, $p < .001$).

To further identify the primary prosodic cues used by NSU in the prosodic encoding of Mandarin intonations, three LDAs were conducted with (1) exclusively global features, (2) exclusively local features, and (3) both global and local features as dependent variables. In Uyghur, the discrimination accuracy was highest when global features were taken into account (96.67%). Thus, NSU relied on the global prosodic features to convey the intonational meanings of statement and question, and the local features were optional. The results of LDA with the highest accuracy for Mandarin intonation are summarized in Table 2.

Table 2: *The prominent prosodic cues of Mandarin intonation.*

Tone	NSM		NSU	
	Prosodic cue	Accuracy	Prosodic cue	Accuracy
T1	Global features	100%	Global + local features	100%
T2	Local features	90%	Global/local features	80%
T3	Global features	80%	Global features	90%
T4	Global + local features	100%	Global features	80%

For T1 sequence, the highest accuracy was achieved by NSU when both global and local features were considered at the same time. In comparison, statement could be distinguished from question by NSM only referring to the global prosodic features. In T2 sequence, based on either global or local features, the accuracy was 80 % for NSU. However, NSM tended only to use pitch-related local features of the last two words and the duration of the initial word. With all T3 syllables, both groups adopted global features only, and pitch-related cues were more important. It indicates that the differences between the two intonations of T3 in the local features were resulted from the interaction between tone and intonation instead of the driving forces for intonational encoding. Regarding the T4 sequence, NSU tended to select global features as the primary reference for the prosodic encoding. However, both global and local features were used by NSM. Less importance of local features in T2 and T4 sequences by NSU aligned with their L1 pattern.

3.2. L1 and L2 influence on English intonation acquisition

3.2.1. Boundary tones

The boundary tones used in Uyghur, Mandarin T1 sequence, and English intonations by NSU, NSM, and NSE were compared in Figure 5. L2 influence was identified by examining whether the same difference in intonation between L1 and L3 existed between L2 and L3.

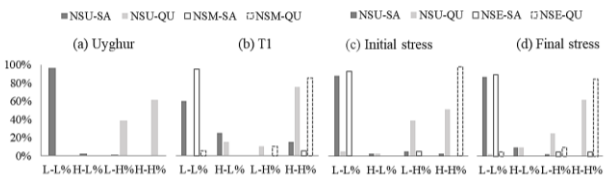


Figure 5: *The inventory and frequencies (%) of IP boundary tones in Uyghur, Mandarin and English intonations.*

Both L-H% and H-H% appeared in the Uyghur interrogative intonation by NSU and Mandarin interrogative intonation by NSM, while L-H% was absent from English intonation with initial stress by NSE. Therefore, it is possible that both L1 and L2 intonations contributed to the additional use of L-H % in English interrogative intonations by NSU when the sentence was composed of initial-stressed words. In declarative sentences, both L-L% and H-L% are used in Uyghur intonation by NSU. In contrast, Mandarin intonation by NSM and English intonations with initial and final stress conditions by NSE mostly ended with L-L%. Thus, when the sentence was composed of final-stressed words, the additional use of H-L% in English intonation by NSU resulted from the influence of L1 declarative intonation.

3.2.2. Prosodic encoding patterns

Figure 6 illustrates the pitch contours of English intonation for declarative and interrogative meanings by NSU and NSE concerning different lexical stress positions.

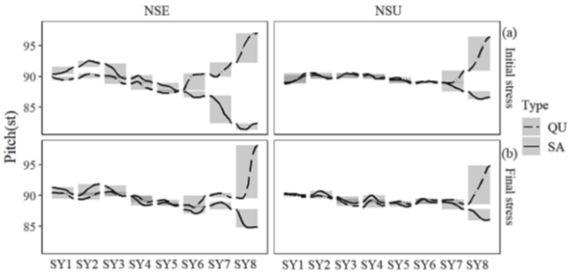


Figure 6: *Pitch curves of English intonation by NSE and NSU with (a) initial and (b) final stressed words.*

In sentences where the lexical stress fell on the initial syllable shown in Figure 6 (a), for NSE, the pitch of the first five syllables in declarative intonation overrode the pitch in interrogative intonation, and the pitch curves intersect at the end of SY5. Differently, from the significant changes in pitch level of SY7 ($t(19) = 2.6$, $p < .05$) and pitch range of SY8 ($t(19) = 4.275$, $p < .001$), the pitch contours of NSU for both declarative and interrogative intonations started to diverge from the beginning of SY7, which was between the positions of division in Uyghur intonation, i.e., SY8, and in L1-English intonations, i.e., SY5. Therefore, the pitch movement patterns of English intonations by NSU was assimilated to L1. In other words, the L1 influence overwhelmed L2 and gave rise to the delay of intonation divergence for statement and question.

As to the English intonations with final-stressed words by NSU, the two intonations were differentiated by the sharp change of pitch in the last syllable. The pitch of interrogative intonation exhibited a noticeable rising trend from the start of SY8, with increasing pitch level ($t(19) = 1.752$, $p < .001$) and pitch range ($t(19) = 2.799$, $p < .05$), which were more similar to both L1-English and L1-Uyghur intonations. It implies that the positive effects from L1 might promoted the learning of English intonation in a native-like approach.

The results of the LDA that achieved the highest discrimination accuracy, signifying the most significant features, are summarized in Table 3.

Table 3: *The prominent prosodic cues of English intonations*.

Stress	NSE		NSU	
	Prosodic cue	Accuracy	Prosodic cue	Accuracy
Initial	Global features	100%	Global features	86.67%
Final	Global features	100%	Global features	73.33%

From the LDA results, the primary prosodic cues to encode English declarative and interrogative intonations were global features for both NSE and NSU, though the overall performance of NSU was not as good as NSE, with a lower accuracy of 86.67% for intonation with initial stress and only 73.33% for intonation with final stress. Specifically, for NSE, the average syllable duration was more useful than overall pitch level. However, the overall pitch cues made more contribution than duration for NSU, which was consistent with Uyghur intonation and indicates the influence from L1.

4. Conclusions

Investigating the performance of Uyghur speakers in Mandarin and English intonation learning within the framework of LILt, this study examines the cross-linguistic influence on L2 and L3 intonation acquisition.

By comparing Mandarin intonations by NSU and NSM with the features of Uyghur intonation as references, the influence of Uyghur in L2 Mandarin intonation acquisition was found in the systemic, frequency, and realizational dimensions, making it similar to L1. Different from Mandarin, no lexical tone system exists in Uyghur. Uyghur speakers had difficulty in encode the interaction between tone and intonation by simultaneous addition [25]. When the original tone shape conflicts with sentence-level prosody, NSU tended to prioritize the phonological and phonetic features of Mandarin intonation with less attention given to lexical tone, that is, they would focus on the contrast between the rising and falling pitch levels and the distinctive difference of boundary tone between the two intonations. For example, in the systemic and frequency dimensions, Uyghur speakers tended to use H-H% in the intonations of T3 and T4 questions. In the realizational dimension, intonations for T2 and T3 questions developed with a gradual rising pitch level and ended with H boundary tone, which was different from the falling intonation characterized by L tonal target as NSM.

Comparing English intonations by NSU and NSE, prominent L1 influence in L3 English intonations by NSU was also identified in the systemic, frequency, and realizational dimensions. Assimilated to Uyghur, when English lexical stress was on the initial syllable, they tended to end with low-rising boundary tone. NSU exhibited a more Uyghur-like pattern in encoding intonations with more focus on the final pitch contrast, which was indicated by the delay of intersect point between pitch curves. The positive transfer also existed when the stress was on the final syllable. L1 intonation could facilitate the prosodic encoding of English intonation in terms of the global manipulation of pitch, making it closer to the L1 norms.

In summary, this study provides evidence supporting the assumptions regarding the cross-linguistic influence proposed by LILt, and its applicability in L3 prosody learning were preliminarily verified by the L1-to-L2, L1-to-L3, and L2-to-L3 influence in intonation learning, which is manifested in the assimilation of L2/L3 intonation to L1. Future research endeavors should consider incorporating data from spontaneous speech, as it can provide a more realistic representation of natural language use. The cross-language standards for the measurements of intonation and bi-directional cross-linguistic influence are also in need of further investigation.

5. Acknowledgements

This research was supported by the Project of Cultural Experts and “Four Batches” of Talents directed by Aijun Li, CASS Institute of Language (2024SYZH001), and the Social Science Foundation of Tianjin, China (Grant No. TJWW19-009).

6. References

[1] D. Bolinger, "Intonation across languages," in *Universals of human language. Vol. 2: Phonology*, Stanford: Stanford University Press, 1978, pp. 471–524.

[2] D. Bolinger, *Intonation and Its Uses: Melody in Grammar and Discourse*. Stanford: Stanford University Press, 1989.

[3] C. Gussenhoven, "Intonation and interpretation: phonetics and phonology," in *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, Apr. 2002, pp. 47–57.

[4] M. Westera, D. Goodhue, C. Gussenhoven, and A. Chen, "Meanings of tones and tunes," in T*he Oxford Handbook of Language Prosody*, Oxford: Oxford University Press, 2020, pp. 443–453.

[5] X. Xie, A. Buxó-Lugo, and C. Kurumada, "Encoding and decoding of meaning through structured variability in intonational speech prosody," *Cognition*, vol. 211, p. 104619, Jun. 2021.

[6] R. Hahn, "Uyghur," in *The Turkic Languages*, London: Routledge, 1998, pp. 379–396.

[7] J. Liang, B. Yi, and Q. Zhang, "An analysis of the stress features of disyllabic words in Uyghur," *Journal of Minzu University of China (Philosophy and Social Sciences Edition)*, vol. 36, no. 2, pp. 115–119, Apr. 2009.

[8] O. Ozçelik, "Stress or intonational prominence? Word accent in Kazakh and Uyghur," *Turkic Languages*, vol. 19, no. 2, pp. 163–192, Jan. 2015.

[9] K. Tursun, "Formal judgment on the Uyghur word stress," *Language and Translation*, no. 4, pp. 38–46, Dec. 2015.

[10] H. Jiang, Y. Liu, and L. Lu, "Acoustics of interrogative intonations in the Uygur language," *Journal of Tsinghua University (Science and Technology)*, no. 48(S1), pp. 635–639, Jan. 2008.

[11] K. Tursun, and B. Wang, "Parallel realization of focus and interrogative meaning in intonation of Uyghur," *Journal of Computer Applications*, vol. 33, no. 3, pp. 784–788, Mar. 2013.

[12] S. Geffen, T. H. Mintz, "Prosodic differences between declaratives and interrogatives in infant-directed speech" *Journal of Child Language*, vol. 44, no. 4, pp. 968–994, Jul. 2017.

[13] F. Liu and Y. Xu, "Question intonation as affected by word stress and focus in English," in *Proceedings of ICPhS 2007 – 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, Aug. 2007, pp. 1189–1192.

[14] M. Safárová, and M. Swerts, "On recognition of declarative questions in English," in *Proceedings of Speech Prosody 2004*, Nara, Japan, Mar. 2004, pp. 313–316.

[15] M. Lin, T*he Experimental Study of Intonation in Mandarin Chinese*. Beijing: China Social Sciences Press, 2012.

[16] J. Yuan, C. Shih, and G. Kochanski, "Comparison of declarative and interrogative intonation in Chinese," in *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, Apr. 2002, pp. 711–714.

[17] J. Liang, and V. J. van Heuven, "Chinese tone and intonation perceived by L1 and L2 listeners," in *Tones and Tunes: Experimental Studies in word and Sentence Prosody*, Berlin: Mouton de Gruyter, 2007, pp. 27–61.

[18] T. Li, H. Feng and Y. Jia, "Prosodic Encoding of Mandarin Chinese Intonation by Uygur Speakers in Declarative and Interrogative Sentences," in *Proceedings of ISCSLP 2022*, Singapore, Dec. 2022, pp. 300–304.

[19] I. Mennen, "Beyond segments: Towards a L2 Intonation Learning Theory," in *Prosody and Language in Contact: L2 Acquisition, Attrition, Languages in Multilingual Situations*, Berlin: Springer, 2015, pp. 171–188.

[20] I. Mennen, U. Reubold, K. Endes, and R. Mayr, "Plasticity of native intonation in the L1 of English migrants to Austria," *Languages*, vol. 7, no. 3, p. 241, Sep. 2022.

[21] C. Draxler, and K. Jänsch, "SpeechRecorder software," Version 6.8.1, 2023. https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/

[22] P. Boersma, and D. Weenink, "Praat: doing phonetics by computer," Version 6.3.19, 2023. http://www.praat.org/

[23] R Core Team, "R: A language and environment for statistical computing," *R Foundation for Statistical Computing*, 2023. https://www.r-project.org/

[24] A. Li, G. Huang, and Z. Li, "Prosodic cues in the interpretation of echo questions in Chinese spoken dialogues," in *Proceedings of 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Lanzhou, China, Sept. 2019, pp. 1056–1061.

[25] Y. R. Chao, "Tone and intonation in Chinese," *Bulletin of the Institute of History and Philology*, vol. 4, no. 2, pp.121–134, 1933.