INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

F0 patterns in Mandarin statements of Mandarin and Cantonese speakers

Yike Yang[1], *Si Chen*[1,2], *Xi Chen*[1]

[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong S.A.R., China

[2]The Hong Kong Polytechnic University-Peking University Research Centre on Chinese Linguistics, Hong Kong S.A.R., China

yi-ke.yang@connect.polyu.hk, sarah.chen@polyu.edu.hk, skye.chen@polyu.edu.hk

Abstract

Cross-linguistic differences of F0 patterns have been found from both monolingual and bilingual speakers. However, previous studies either worked on intonation languages or compared an intonation language with a tone language. It still remains unknown whether there are F0 differences in bilingual speakers of tone languages. This study compared second language (L2) Mandarin with Cantonese and first language (L1) Mandarin, to test whether the L2 speakers of Mandarin have acquired the F0 patterns of Mandarin and whether there are influences from their L1 Cantonese. Different F0 measurements (including maximum F0, minimum F0, mean F0 and F0 range) were examined with linear mixed-effects models. Cantonese and Mandarin showed different F0 patterns, the source of which still requires further investigation. The L2 Mandarin data resembled the F0 patterns of Cantonese and were different from L1 Mandarin, for which we provided different explanations: assimilation of L1 Cantonese and L2 Mandarin, the negative transfer from native Cantonese, and similarities in the nature of tone languages. Suggestions for testing these assumptions are proposed. Lastly, our data provided conflicting results concerning the role of gender in F0 pattern realisation.

Index Terms: speech production, second language, F0 pattern, Mandarin, Cantonese

1. Introduction

Languages differ in their F0 profiles, and studies have been conducted to test how speakers of various languages make use of F0-related cues, especially mean F0 and F0 range [1]–[6]. For example, Mandarin speakers were found to have higher mean F0 and more F0 variation than American English speakers in read speech, and the divergence was attributed to the different intonation patterns of a tone language and a stress language [1]. Even within the system of intonation languages, English female speakers showed a higher F0 level and a wider F0 span than German female speakers [4]. Several sources of cross-linguistic F0 differences have been proposed, such as differences in the intonation of languages, divergence in cultural and social norms and differences in physiology of speakers [2], but no consensus has been reached on this issue. Also, such divergence in F0 profiles is observed for the two languages of simultaneous bilinguals. In Japanese-English bilinguals' read speech, Japanese had higher sentence-initial F0, higher first and second peaks as well as steeper declination lines than English [3]. Moreover, there might be a gender effect that interacts with the F0 profiles of languages, but the mechanism behind remains unclear [2], [5]. For instance, consistent cross-linguistic differences in F0 range were observed in Welsh-English female bilinguals but not in male bilinguals [2].

However, there is little information about F0 patterns of the two languages in late second language (L2) learners, or sequential bilinguals, who usually start to learn their L2 after complete acquisition of their first language (L1) and are associated with accented speech in their L2 [7]. Work into this direction can deepen our understanding on issues of L2 speech and foreign accent, which will shed light on the postulations from models such as Speech Learning Model (SLM) [8], [9] and Perceptual Assimilation Model (PAM) [10]. Although such models were originally proposed to account for the learning of L2 segments, it is possible to borrow the ideas into the investigations on L2 prosody. Nguyễn [11] examined the production of English by English and Vietnamese speakers. The results showed that advanced Vietnamese speakers and native English speakers shared similar F0 patterns, which had greater F0 variation than beginners. The beginners' English, on the other hand, resembled the F0 patterns of Vietnamese, and this was interpreted as being transferred from Vietnamese.

To the best of our knowledge, no thorough investigation has been conducted on tone language pairs, so this study attempts to fill this gap with data from speakers of two closely related tone languages, namely, Mandarin and Cantonese.

Mandarin and Cantonese are tone languages from the Sino-Tibetan language family, but they have different phonological systems and are not mutually intelligible [12]. There are four lexical tones in Mandarin, among which Tone 1 (represented as 55) is the high-level tone and the remaining ones are contour tones [13]. In Cantonese, there are six lexical tones and three of them are level tones: high-level Tone 1 (55), mid-level Tone 3 (33) and low-level Tone 6 (22) [14]. To accommodate more tones, Cantonese is shown to have a wider F0 space than Mandarin (e.g. F0 range of male speakers for the six Cantonese tones: 80-170 Hz [15]; F0 range of male speakers for the four Mandarin tones: 90-140 Hz [16]). Despite the phonological differences, Chinese characters are used in the orthography of both languages.

The current study first compared the F0 patterns of Mandarin and Cantonese, and then compared L2 Mandarin with Cantonese and L1 Mandarin, to test whether the L2 speakers of Mandarin have acquired the F0 patterns of Mandarin and whether there are influences from their L1 Cantonese. Also, the effect of gender was taken into account in this study. The F0 profiles tested in this study included maximum F0, minimum F0, mean F0 and F0 range.

http://dx.doi.org/10.21437/Interspeech.2020-2549

2. Methods

2.1. Participants

Eleven native speakers of Mandarin (six females, five males; aged: 24.72 ± 4.39) and 12 native speakers of Hong Kong Cantonese (six females, six males; aged: 20.41 ± 2.97) participated in a production experiment at the Speech and Language Sciences Laboratory of the Hong Kong Polytechnic University. The Mandarin participants started to speak Mandarin from birth and had spent most of their lives in Mandarin-speaking regions. The Cantonese participants speak Cantonese as the native and dominant language and also learned to speak Mandarin as a second language at an average age of three years and nine months old. To assess their language profile of Cantonese and Mandarin, the Cantonese speakers completed a language background questionnaire, which was an adapted version of Bilingual Language Profile [17]. Information concerning language history, language use, language proficiency and language attitudes was collected and converted to both module and global scores, the results of which suggested that the Cantonese speakers are Cantonese-dominant but are also fluent in Mandarin. No participants received formal musical training, and none reported any history of speaking, hearing or language difficulty.

2.2. Materials

The experiments were designed for a larger project on Mandarin and Cantonese prosody and part of the Mandarin data has been reported in [18]. Short subject-verb-object (SVO) statements were used as the test sentences. To make the data of Cantonese and Mandarin comparable, exactly the same syntactic structure was adopted for Cantonese and Mandarin stimuli, with seven syllables in each sentence. A determiner and a classifier (DET and CL; both monosyllabic) were added to the beginning of each sentence so that the stimuli make more sense. We were more interested in the global F0 patterns than the tonal effect, so for the remaining five syllables, only Tone 1 in both languages was used in the sentences (it is very difficult, if not impossible, to control the tones of the determiners and classifiers in both languages). Sentences (1a) and (1b) are examples of the Mandarin and Cantonese target sentences, respectively.

(1) a. *na4 wei4 yi1sheng1 he1 ka1fei1*
DET CL doctor drink coffee
'The doctor drinks coffee.'

b. *go2 di1 can1cik1 zong1 faa1dang1*
DET CL relative install lantern
'The relatives install lanterns.'

The Mandarin speakers recorded six target sentences in Mandarin, and the data were then labelled as 'Man_L1' (L1 Mandarin). The Cantonese speakers recorded four target sentences in Mandarin and another four in Cantonese, which were labelled as 'Man_L2' (L2 Mandarin) and 'Can' (L1 Cantonese), respectively. Three repetitions were collected for each target sentence. In total, there were 198 tokens (6 sentences * 11 speakers * 3 repetitions) in the 'Man_L1' data group, 144 tokens (4 sentences * 12 speakers * 3 repetitions) in the 'Man_L2' data group, and 144 tokens (4 sentences * 12 speakers * 3 repetitions) in the 'Can' data group.

2.3. Data collection

The sentences, written in Chinese characters, were divided into different blocks and randomly presented on a computer screen in E-Prime 2.0 [19]. The sentences were elicited as answers to questions the experimenter asked. This procedure has the following advantages: 1) this semi-simultaneous elicitation made the data collection more natural than read speech; and 2) with such method, the experiment also managed to control the recording materials and made them comparable across speakers and languages. The dialogues were recorded at a sampling rate of 44,100 Hz in Audacity [20] on another computer. Only the answers were further segmented and analysed.

This project has been approved by the Human Subjects Ethics Sub-committee (HSESC) of the Hong Kong Polytechnic University (Reference #: HSEARS20190102001). All participants gave their written informed consent prior to the recording sessions.

2.4. Data analysis

Each syllable of the target sentences was manually segmented in Praat [21], and the maximum F0, minimum F0, mean F0 and F0 range were extracted using the ProsodyPro Praat script [22]. The F0 values, originally measured in Hz, were converted to semitones (st) individually, with mean F0 of each speaker as reference [23].

As mentioned in Section 2.2, each sentence starts with a determiner and a classifier, which are not in Tone 1. In our analysis, we first created two subsets of data, one with the complete sentences and another with the remaining five Tone 1 syllables only, and analysed each subset separately with linear mixed-effects models using the 'lme4' package [24] in R [25], [26]. According to the models, there was no significant difference between the two subsets of data, so in the results, we presented the sentences with five Tone 1 syllables only. Because we wanted to examine the F0 patterns at both the syllable and utterance levels, we further divided the subsets into two groups. The F0 values were our dependent variable. *Group*, *Gender* and their interaction were the fixed effects. *Sentence*, *Repetition*, *Speaker* and *Syllable* (for the syllable level only) were included as the random effects. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. If the effect of *Gender* was significant, we would separate the data into two subsets of female and male speakers. Also, to further test the difference among the groups, pairwise comparisons of different language groups were conducted. The figures were plotted with the 'ggplot2' package [27].

3. Results

3.1. F0 patterns at the utterance level

We first fit models for the F0 patterns at the utterance level and plotted the results in Figure 1. There were main effects of *Group* for minimum F0, F0 range and mean F0 (ps < .001) and main effects of *Gender* for the maximum F0 and minimum F0 (ps < .002). There was no interaction of *Group* and *Gender* across all the variables.

For female speakers, main effects of *Group* were found for minimum F0, F0 range and mean F0 (ps < .005). Pairwise comparisons suggested that the L1 Mandarin data had lower maximum F0 than the Cantonese data (p = .049) and higher minimum F0 and mean F0 than the Cantonese and L2 Mandarin data ($ps < .006$). The Mandarin female speakers also had a smaller F0 range than the other groups ($ps < .003$). No differences were found between the Cantonese and L2 Mandarin of the female speakers. For male speakers, main effects of *Group* were found for minimum F0, F0 range and mean F0 ($ps < .019$). The Mandarin male speakers showed higher minimum F0 and mean F0 ($ps < .001$) as well as a smaller F0 range ($ps < .02$) than the other two groups. There was no difference between the Cantonese and L2 Mandarin data of the male speakers.

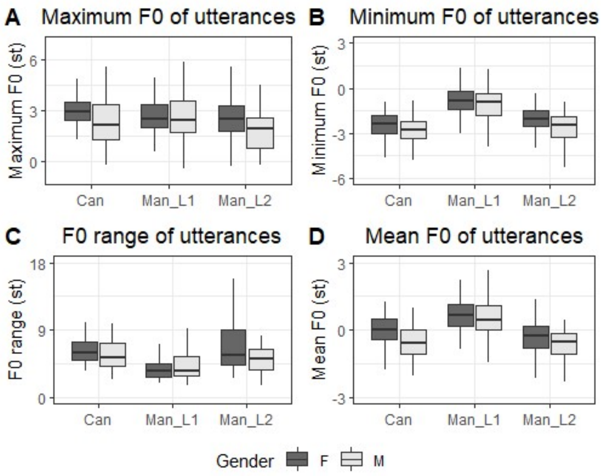

Figure 1: *F0 values of the utterances.*

At the utterance level, the variables *Group* and *Gender* affected the F0 values. A closer examination suggested that the L1 Mandarin data differed from the Cantonese and L2 Mandarin data in having higher minimum F0 and mean F0 and a smaller F0 range, and there was no difference in the F0 profiles of the Cantonese and L2 Mandarin data. Similar patterns were found in female and male speakers.

3.2. F0 patterns at the syllable level

We then employed the linear mixed-effects models to analyse the F0 patterns at the syllable level, the results of which are presented in Figure 2. There were main effects of *Group* for the maximum F0, minimum F0, F0 range and mean F0 ($ps < .001$) and main effects of *Gender* for the maximum F0, minimum F0 and mean F0 ($ps < .012$). Interactions of *Group* and *Gender* were also found for F0 range and mean F0 ($ps < .043$).

For female speakers, there were effects of *Group* for the minimum F0, F0 range and mean F0 ($ps < .001$). The L1 Mandarin data exhibited higher maximum F0 than the L2 Mandarin data (p = .036) and higher minimum F0 and mean F0 than the Cantonese and L2 Mandarin data ($ps < .001$). Also, there was a smaller F0 range for the L1 Mandarin speakers ($ps < .001$). Again, no difference was found in the Cantonese and L2 Mandarin of the female speakers. For male speakers, significant effects of *Group* were also found for the minimum F0, F0 range and mean F0 ($ps < .001$). The L1 Mandarin data were highest in minimum F0 and mean F0 ($ps < .001$), and the L2 Mandarin data were lower than the Cantonese data in minimum F0 (p = .049). In terms of F0 range, L1 Mandarin was lower than L2 Mandarin (p = .001) and Cantonese (p = .057), and L2 Mandarin was higher than Cantonese (p = .009).

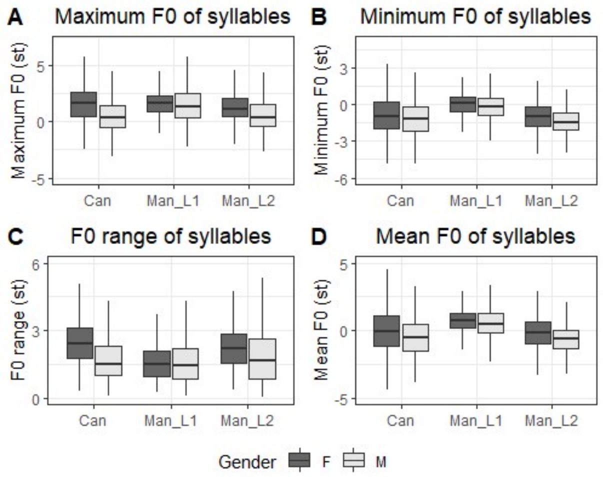

Figure 2: *F0 values of the syllables.*

At the syllable level, the variables *Group* and *Gender* also affected the F0 values. The female data at the syllable level resembled those at the utterance level, where the L2 Mandarin and Cantonese data showed very similar patterns and the L1 Mandarin data revealed systematic differences with the other two groups. For male speakers, the three groups of data showed noticeable differences, although the L1 Mandarin data had the highest mean F0 and the smallest F0 range, consistent with data from female speakers.

4. Discussion

The first aim of the study was to examine the F0 profiles of Mandarin and Cantonese. Across different levels (syllable and utterance) and genders, the general patterns identified were: 1) Mandarin and Cantonese speakers had comparable maximum F0 in their native languages, suggesting that they share similar upper limit in their use of F0 (at least for the normalised data according to our analysis); 2) Mandarin speakers showed higher minimum F0 and consequently a smaller F0 range than Cantonese speakers, which falls well in line with previous findings that Cantonese speakers have a wider F0 span than Mandarin speakers [15], [16]; and 3) Mandarin speakers had higher mean F0 than Cantonese speakers.

There are six lexical tones in Cantonese, and three of them are level tones. Among the four lexical tones in Mandarin, there is only one level tone. It is thus reasonable for Cantonese to have a larger F0 space than Mandarin to maintain the tonal contrasts. Yet it is unclear why Cantonese has lower mean F0 than Mandarin. Further investigations are required as to whether the difference is due to physiological constraints of speakers, the relatively small sample size of this study, or language-specific features of Mandarin and Cantonese.

Our second aim was to investigate the L2 Mandarin data produced by Cantonese speakers and compare them with native Mandarin and Cantonese. The data revealed that the L2 Mandarin resembled Cantonese in our measurements and were strikingly different from L1 Mandarin. Specifically, compared with the L2 Mandarin, the L1 Mandarin had a smaller F0 range but higher mean F0. Given the fact that all Cantonese learners speak fluent Mandarin, it is surprising to find such huge divergence from the L1 and L2 Mandarin data. In a recent study on Mandarin tone sandhi production [28], Cantonese speakers also had lower F0 values than Mandarin speakers under some tonal combinations. This consistent pattern might be a result of the Cantonese learners' perceptual mapping of Cantonese and Mandarin tones [29], because there are differences in the tonal systems of these two languages.

The category assimilation hypothesis (CAH) from SLM may provide another possible explanation. According to SLM, there is a common phonological space for storing L1 and L2 sounds [8]. The CAH suggests that an L2 sound perceived similar to an L1 sound does not form a new category, and is understood as a variant of the L1 sound at an allophonic level. Such mapping eventually gives rise to a new merged category in mental representation of a bilingual speaker. In our test, the Cantonese speakers may have stored the prosodic features of F0 in L1 Cantonese and L2 Mandarin in the common phonological space and do not specifically differentiate them. As a result, the features of F0 in L1 and L2 have undergone assimilation processes and become more similar to each other. To test whether this is the case, two directions can be considered. The participants from the current study are fluent L2 speakers of Mandarin, so subsequent studies may invite learners with much less exposure to Mandarin and at lower proficiency levels (beginner and intermediate levels). Also, Mandarin sentences varying in F0 values can be used as the stimuli for perceptual experiments to test Cantonese speakers' sensitivity to Mandarin F0 levels.

However, one might argue that the similarities in Cantonese speakers' Mandarin and Cantonese actually reflect their non-nativeness in Mandarin, although they speak fluent Mandarin. This claim seems convincing as L2 speech learning, especially L2 prosody, is very challenging, and the attainment of native pronunciation is unlikely for late L2 learners [30]. It may be the case that our participants have not fully acquired the features of Mandarin and still have trouble with the production of Mandarin prosody. As a result, their L2 Mandarin production reveals a negative transfer from their native Cantonese. However, only when Cantonese learners of Mandarin at different proficiency levels are tested (as suggested in the previous paragraph), can we verify whether this claim holds.

Also, the similar F0 profiles in Cantonese speakers' Cantonese and Mandarin observed in our study lend some support to the physiology-based claim for cross-linguistic F0 differences. However, special caution must be taken before we can reach a conclusion because previous studies showed differences in bilinguals' language pairs (e.g. English and Korean [5], Welsh and English [2]). Unlike previous studies, our data were collected from speakers of two tone languages, and we thus provided new data regarding cross-linguistic differences of F0 patterns. For tone languages, each syllable determines its local F0 contour, which interacts with intonation in the realisation of the global F0 contour in an utterance. There are remaining issues to be examined. First, as stated before, it is yet unknown whether of Mandarin and Cantonese share similar intonation patterns. Second, because individual differences are common in speech production [31], it is necessary to test the F0 patterns of the two tone languages within the same speaker and across different speakers. Third, more speakers with diverse language backgrounds (especially Cantonese-Mandarin bilinguals that are more Mandarin-dominant) are needed to address this issue.

The third aim of the study was to examine the role of gender in cross-linguistic F0 formation. The L1 Mandarin was distinct from the L2 Mandarin and Cantonese regardless of gender, but the effect of gender was found in Cantonese speakers' Cantonese and L2 Mandarin. For female speakers, there was no difference between the Cantonese and L2 Mandarin data at both the utterance and syllable levels. For male speakers, although the patterns of Cantonese and L2 Mandarin were similar at the utterance level, the two sets of data diverged from each other at the syllable level. Previous investigations on this issue provided conflicting results. While some studies suggested that female bilinguals tended to switch the F0 patterns consistently between the two languages [2], others showed no effect of gender on the F0 patterns [3]. Contradicting previous findings, our results complicated the issue of gender effect, which remains to be explored.

To conclude, Cantonese and Mandarin showed different F0 patterns, the source of which still requires further investigation. The L2 Mandarin data resembled the F0 patterns of Cantonese and were different from L1 Mandarin, for which we provided different explanations: assimilation of L1 Cantonese and L2 Mandarin, negative transfer from native Cantonese, and similarities in the nature of tone languages. Suggestions for testing these assumptions are proposed. Lastly, our data provided conflicting results concerning the role of gender in F0 pattern realisation.

5. Acknowledgements

The authors thank all the informants for their participation in this project and the anonymous reviewers for their comments on this paper.

6. References

[1] S. J. Eady, "Differences in the F0 Patterns of Speech: Tone Language Versus Stress Language," *Lang. Speech*, vol. 25, no. 1, pp. 29–42, 1982.

[2] M. Ordin and I. Mennen, "Cross-Linguistic Differences in Bilinguals' Fundamental Frequency Ranges," *J. Speech Lang. Hear. Res.*, vol. 60, no. 6, p. 1493, 2017.

[3] C. Graham, "Fundamental Frequency Range in Japanese and English: The Case of Simultaneous Bilinguals," *Phonetica*, vol. 71, no. 4, pp. 271–295, 2014.

[4] I. Mennen, F. Schaeffler, and G. Docherty, "Cross-language differences in fundamental frequency range: A comparison of English and German," *J. Acoust. Soc. Am.*, vol. 131, no. 3, pp. 2249–2260, 2012.

[5] A. Cheng, "Cross-linguistic f0 differences in bilingual speakers of English and Korean," *J. Acoust. Soc. Am.*, vol. 147, pp. EL67–EL73, 2020.

[6] P. Keating and G. Kuo, "Comparison of speaking fundamental frequency in English and Mandarin," *J. Acoust. Soc. Am.*, vol. 132, no. 2, pp. 1050–1060, 2012.

[7] M. S. Schmid and H. Hopp, "Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects," *Lang. Test.*, vol. 31, no. 3, pp. 367–388, 2014.

[8] J. E. Flege, "Second Language Speech Learning: Theory, Findings, and Problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language research*, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 233–277.

[9] J. E. Flege, "Interactions between the native and second-language phonetic systems," in *An integrated view of language development: Papers in honor of Henning Wode*, T. Piske, A. Rohde, and P. Burmeister, Eds. Trier: Wissenschaftlicher Verlag, 2002, pp. 217–244.

[10] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," in

Second language speech learning: The role of language experience in speech perception and production, M. J. Munro and O.-S. Bohn, Eds. Amsterdam: John Benjamins, 2007, pp. 13–34.
[11] A. T. T. Nguyễn, "F0 patterns of tone versus non-tone languages: The case of Vietnamese speakers of English," *Second Lang. Res.*, vol. 36, no. 1, pp. 97–121, 2020.
[12] X. Zhang, "Dialect MT: A case study between Cantonese and Mandarin," in *Proc. COLING 1998*, 1998, vol. 2, pp. 1460–1464.
[13] Y. R. Chao, *Mandarin Primer*. Cambridge: Harvard University Press, 1948.
[14] R. S. Bauer and P. K. Benedict, *Modern Cantonese Phonology*. Berlin: Walter de Gruyter, 1997.
[15] A. L. Francis, V. Ciocca, L. Ma, and K. Fenn, "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers," *J. Phon.*, vol. 36, no. 2, pp. 268–294, 2008.
[16] Y. Xu, "Contextual tonal variations in Mandarin," *J. Phon.*, vol. 25, no. 1, pp. 61–83, 1997.
[17] D. Birdsong, L. M. Gertken, and M. Amengual, "Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism," *COERLL, University of Texas at Austin*, 2012. https://sites.la.utexas.edu/bilingual/.
[18] Y. Yang and S. Chen, "Revisiting focus production in Mandarin Chinese: Some preliminary findings," in *Proc. Speech Prosody 2020*, 2020, pp. 260–264.
[19] W. Schneider, A. Eschman, and A. Zuccolotto, *E-Prime User's Guide*. Pittsburgh: Psychological Software Tools Inc, 2012.
[20] Audacity Team, "Audacity(R): Free Audio Editor and Recorder." 2019.
[21] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." 2015.
[22] Y. Xu, "ProsodyPro - A tool for large-scale systematic prosody analysis," in *Proc. TRASP'2013*, 2013, pp. 7–10.
[23] F. Nolan, "Intonational equivalence: an experimental evaluation of pitch scales," in *Proc. ICPhS 2003*, 2003, pp. 771–774.
[24] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, 2015.
[25] R Core Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2018.
[26] RStudio Team, "RStudio: Integrated Development for R." RStudio, Inc., Boston, MA, 2016.
[27] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer, 2016.
[28] S. Chen, Y. He, R. Wayland, Y. Yang, B. Li, and C. W. Yuen, "Mechanisms of tone sandhi rule application by tonal and non-tonal non-native speakers," *Speech Commun.*, vol. 115, pp. 67–77, 2019.
[29] Y. C. Hao, "Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers," *J. Phon.*, vol. 40, no. 2, pp. 269–279, 2012.
[30] D. Singleton, "The Critical Period Hypothesis: A coat of many colours," *IRAL - Int. Rev. Appl. Linguist. Lang. Teach.*, vol. 43, no. 4, pp. 269–285, 2005.
[31] Y. Yang and S. Chen, "Individual differences in Mandarin focus production," in *Proc. ExLing 2020*, 2020.