



Acoustic Features of Mandarin Tone Production in Noise: A Comparison Between Chinese Native Speakers and Korean L2 Learners

Jinxin Ji^{1,2}; Yiyi Hu¹; Xiaohu Yang^{2*}; Gang Peng^{1*}

¹ Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR

² Department of English, Tongji University, China

gang.peng@polyu.edu.hk, 19026@tongji.edu.cn

Abstract

Research on Lombard speech produced by Chinese learners remains underdeveloped. This study investigates how native Chinese Mandarin speakers and Korean second language (L2) learners adjust the production of lexical tones under noise conditions. Fifty-two participants (26 Chinese native speakers and 26 higher-level Korean L2 learners) produced 28 syllables under three conditions: babble noise 80 dB SPL, white noise 80 dB SPL, and quiet. Acoustic features, including intensity, duration and F0 contours were analysed. Results indicated that Korean L2 learners significantly increased both intensity and duration across all four tones in the two noise conditions, whereas Chinese native speakers significantly increased durations only for Tone 3 and Tone 4 under white noise. Both groups adjusted F0 contours for all tones in noise, with a notable group difference in the F0 contour of Tone 3. These findings were discussed in association with feedback control theory in second language learners, showing that L2 learners rely more heavily on feedback control and are more detrimentally affected by noise conditions than native speakers.

Index Terms: second language production, lexical tone, speech production, noise, Lombard effect

1. Introduction

Native speakers typically increase intensity, adjust fundamental frequency (F0), and lengthen word durations in noisy environment compared with quiet environment [1]. This involuntary tendency is known as the Lombard effect [2]. Previous research has further revealed that white noise having a more significant effect than other noise types [3].

Several studies have examined English as a second language (L2) learners' speech production in noisy conditions. These studies have shown that learners exhibited the Lombard effect in their L2 [4, 5, 6]. Differences in adjustments across acoustic features were observed between first language (L1) and L2 processing. Chinese learners of English showed a greater increase in intensity in L2 English Lombard speech than in L1 [7, 8], and similar results were found for Japanese learners of English [9]. Studies among Dutch native learners of English revealed that compared with native English speakers, these L2 learners exhibited a smaller increase in F0 range [10] and in median F0 [11] when producing English sentences in Lombard speech. A study of Chinese L2 learners showed that they decreased F0 in L2 English and increased F0 in L1 in noise compared with quiet [5].

Research on the Lombard effect in L2 speech in languages other than English is underdeveloped. Our previous study [12] investigated how Korean learners produce Mandarin tones in quiet and noise conditions. The results indicated that both Korean and Chinese speakers exhibited similar increases in intensity and duration under noisy conditions compared with quiet conditions. Chinese learners and Korean L2 speakers exhibited similar F0 contours for Tone 2 and Tone 3 in quiet, while showing different adjustments in noise conditions. However, this study has three limitations. Firstly, the sample size of the speakers was relatively small, including only 5 native speakers and 5 L2 learners, which might limit generalizability. Secondly, the analysis used ANOVA, which did not account for the random effects of subjects, despite previous research showing significant individual differences in L2 speech production among learners [13]. Thirdly, the Korean participants had varying proficiency levels, making it difficult to draw clear conclusions from the results.

A related theory that explains the differences between native speakers and L2 learners is the feedback control theory in L2 production [7, 8]. Speech motor control models involve the coordination of feedforward control and feedback control [14]. Feedforward control executes the motor commands for targeted speech movements, while feedback control operates through speech perception, monitoring speech production, and issuing corrective commands for speech errors. Adult native speakers rely highly on feedforward control, since their speech errors are fewer than L2 learners. The reliance on auditory feedback control is greater for L2 learners, who are more prone to speech errors. Therefore, the speech production of L2 speakers is more influenced by noise due to the detrimental effects of noise on speech perception, which in turn affects feedback control.

In the present study, we investigated how higher-level Korean native learners of L2 Chinese produce Mandarin tones under noise conditions, using a larger sample and analysing the data with mixed-effects models. Standard Korean is a non-tonal language, while Mandarin is a tonal language with four lexical tones. Previous studies have shown that Korean learners of Chinese have difficulties with F0 contours for Mandarin tones, especially for Tone 3 [15, 16]. However, most of these studies were conducted under quiet condition. The noise level was set to a sound pressure level of 80 dB, as previous research [12] and our preliminary experiment showed that Chinese and Korean learners began to adjust their speech at this intensity.

2. Method

2.1. Participants

A total of 52 participants took part in the experiment, including 26 native Mandarin Chinese speakers (13 males, 13 females, Mean Age = 25.62 years, $SD = 3.02$) and 26 Korean learners of Chinese (13 males, 13 females, Mean Age = 25.49 years, $SD = 2.88$). The native Mandarin speakers came from northern China and were undergraduate and graduate students. The Korean participants were undergraduate and graduate students recruited from universities in Shanghai. None of the Korean participants' native languages was Busan Korean. The language of instruction in their university programs was Mandarin, and all of them had passed the HSK Level 4 test before coming to China as the program entrance requirement. On average, the age of acquisition (AOA) was 18 years ($SD = 5.51$). All Korean L2 learners had lived in China for over one year, with an average length of residence (LOR) being 3.94 years ($SD = 3.79$). According to their reports, they used Chinese 72.5% of the time in their studies ($SD = 13.69$) and used Chinese 63.89% of the time in their daily lives ($SD = 19.70$).

2.2. Materials

The materials consisted of all combinations of 7 different vowels and 4 tones, totalling 28 monosyllabic words (see Table 1). These 7 vowels were /ɿ, a, i, u, y, ɿ, ʌ/. White noise was continuous, broadband noise with a wide frequency range from 125 to 8000 Hz. The babble noise consisted of two individuals simultaneously reading two distinct sets of materials, with the male focusing on world geography and the female on climate change. The two noise conditions were played at 80 dB SPL.

Table 1: *Materials.*

Tone	[kɿ]	[a]	[di]	[du]	[tɕy]	[tsʰ]	[tsʰ]
Tone 1	哥 ge1 brother	差 da1 brother	滴 di1 drip	督 du1 supervise	居 ju1 reside	吃 chi1 eat	吃 chi1 eat
Tone 2	格 ge2 gridwork	答 da2 answer	笛 di2 flute	读 du2 read	局 ju2 bureau	迟 chi2 late	迟 chi2 benevolent
Tone 3	哥 ge3 kudzu	打 da3 hit	底 di3 bottom	督 du3 bet	举 ju3 raise	耻 chi3 shame	此 ci3 this
Tone 4	个 ge4 a	大 da4 big	弟 di4 brother	肚 du4 belly	句 ju4 sentence	赤 chi4 red	刺 ci4 stab

2.3. Procedure

The recordings were conducted in a soundproof room, using an external microphone positioned approximately 10 cm from the participants' mouths. Participants attended the experiment twice, with the three conditions presented in random order across three separate blocks. During the two noise conditions, the noise sounds were played through Sennheiser HD 280 PRO headphones, adjusted to 80 dB using an artificial ear. Within each block, 28 syllables (presented with tone numbers, pinyin, and characters) were randomly displayed on the screen in the carrier sentence “我读_____” (“I read_____”) using E-Prime 3. Participants were instructed to read the sentence three times.

2.4. Data Collection and Analysis

Acoustic data were extracted from the second repetition of each syllable, as it represented the most stable pronunciation, resulting in a total of 28 syllables \times 52 participants \times 3 conditions = 4,368 tokens. The annotations were carried out by two native Mandarin-speaking postgraduate students in linguistics, both with phonetic training, using Praat. Acoustic

data were measured across three dimensions: duration, mean intensity, and F0 contours.

The F0 values were extracted at ten equidistant points using the script [17] and were z-score normalized based on the mean pitch for each individual speaker, using the formula: Normalized pitch = (Observed pitch - Mean pitch) / Standard deviation of pitch [18], following previous studies [19, 20]. Linear mixed-effects models (LMMs) in R were used to analyse intensity and duration, employing the lme4 package [21]. Generalized additive mixed-effects models (GAMMs) were used to model F0 contours for its advantage over other time-series analyses in modelling complex nonlinear relationships [20], using the mgcv package [22]. For each acoustic feature, models were constructed for each tone. The dependent variable was the acoustic feature, with fixed effects including group (Korean and Chinese) and condition (quiet, white noise, babble noise). The random effects structure included subject and item. Bonferroni post-hoc tests were conducted for pairwise comparisons when significant main effects or interactions were found, using the emmeans package [23]. Plots were generated using the ggplot2 package [24].

3. Results

3.1. Mean Intensity

Table 2 presents the results of the models for mean intensity. Figure 1 shows the intensity of all four tones for both groups. Significant *group* \times *condition* interactions were observed for all four tone models. Bonferroni post-hoc tests for the interactions revealed that Chinese speakers did not significantly adjust their mean intensity in the two noise conditions compared with the quiet condition for any of the four tones (all $ps > 0.05$). In contrast, Korean L2 learners showed significantly higher mean intensity in the two noise conditions compared with the quiet condition (all $ps < 0.01$). These results suggested that while Chinese speakers maintained a consistent intensity across all conditions, Korean L2 learners increased their voice intensity in the two noise conditions.

Table 2: *LMMs Results for Intensity.*

Term	Tone											
	Tone 1			Tone 2			Tone 3			Tone 4		
	χ^2	df	p	χ^2	df	p	χ^2	df	p	χ^2	df	p
Intercept	18874	1	<0.001***	12956	1	<0.001***	7314	1	<0.001***	8512	1	<0.001***
Group	4.82	1	0.028*	7.10	1	0.007**	1.83	1	0.18	4.25	1	0.039*
Noise	8.62	2	0.013*	18.05	2	<0.001***	4.94	2	0.09	6.02	2	0.049*
Group: Noise	106.60	2	<0.001***	64.75	2	<0.001***	43.20	2	<0.001***	62.32	2	<0.001***

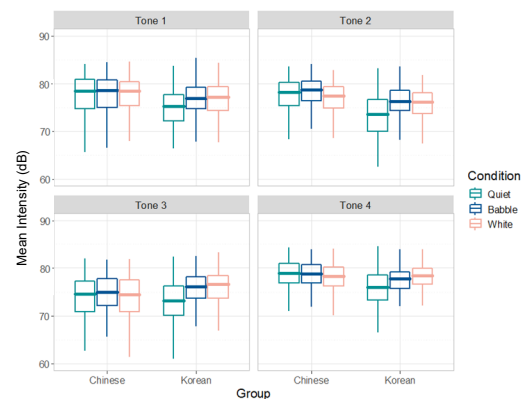


Figure 1: *Intensity for Speakers.*

3.2. Duration

Table 3 presents the results of the duration models. Figure 2 shows the duration of the four tones for both groups. Significant *group* × *condition* interactions were observed in the models for Tone 1, Tone 2, and Tone 3. For Tone 1 and Tone 2, Bonferroni post-hoc tests for the interactions revealed that Chinese speakers did not significantly adjust their duration in the noise conditions compared to the quiet condition (all *ps* > 0.05). In contrast, Korean L2 learners exhibited significantly longer durations in the two noise conditions compared with the quiet condition (all *ps* < 0.05).

For Tone 3, Bonferroni post-hoc tests for the interactions revealed that both Korean and Chinese speakers significantly adjusted their duration in the white noise condition compared with the quiet condition (all *ps* < 0.05). Only Korean speakers showed significant adjustments under the babble noise condition compared with the quiet condition ($\beta = 20.94$, $SE = 6.42$, $t = 3.26$, $p = 0.017$), while Chinese speakers maintained similar duration under these two conditions. In addition, both groups showed significant adjustments under the white noise condition compared with the babble noise condition (all *ps* < 0.05).

For Tone 4, significant main effects of group and condition were observed. Bonferroni post-hoc tests revealed that Korean L2 learners had significantly longer durations than Chinese speakers ($\beta = -15.30$, $SE = 7.07$, $t = -2.17$, $p = 0.035$). Both noise conditions showed significantly longer durations than the quiet condition (all *ps* < 0.001), and the duration under the white noise condition was significantly longer than under the babble noise condition ($\beta = -8.24$, $SE = 2.34$, $t = -3.52$, $p = 0.001$). Further analysis revealed that both Korean and Chinese speakers significantly adjusted their duration in white noise condition compared with the quiet condition (all *ps* < 0.05).

Table 3: LMMs Results for Duration.

Term	Tone											
	Tone 1			Tone 2			Tone 3			Tone 4		
	χ^2	<i>df</i>	<i>p</i>	χ^2	<i>df</i>	<i>p</i>	χ^2	<i>df</i>	<i>p</i>	χ^2	<i>df</i>	<i>p</i>
Intercept	363.08	1	<0.001***	445.91	1	<0.001***	342.37	1	<0.001***	507.22	1	<0.001***
Group	28.00	1	<0.001***	15.21	1	<0.001***	4.11	1	0.043*	5.14	1	0.023*
Noise	1.47	2	0.48	1.83	2	0.40	84.42	2	<0.001***	25.63	2	<0.001***
Group: Noise	13.46	2	0.001***	20.91	2	<0.001***	8.82	2	0.012*	0.69	2	0.71

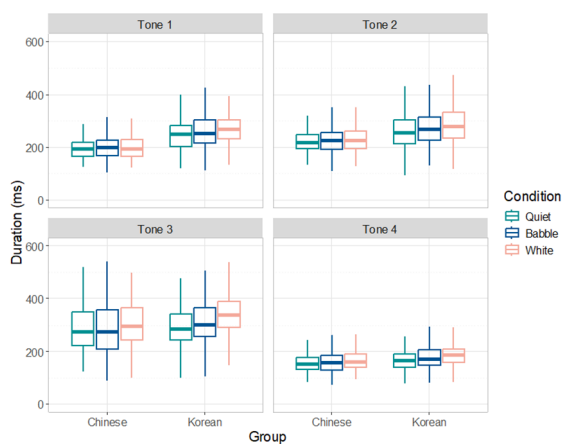


Figure 2: Duration for Speakers.

3.3. F0 Contour

Table 4 presents the results for the F0 contour models. Figure 3 shows the F0 contours for four tones. Significant *group* ×

condition interactions were found for each of the four tone models. For Tone 1, Bonferroni post-hoc tests showed that Chinese and Korean participants made significant adjustments under the two noise conditions compared with the quiet condition (all *ps* < 0.001). Both Chinese native speakers and Korean L2 learners exhibited significant differences between the white noise and babble noise conditions, with the F0 contour in the white noise being significantly higher than in the babble noise (all *ps* < 0.001). Furthermore, there were no significant differences between groups in either the noise or quiet conditions.

For Tone 2, Bonferroni post-hoc tests showed that Chinese and Korean participants made significant adjustments under the two noise conditions compared with the quiet condition (all *ps* < 0.001). Both Chinese native speakers and Korean L2 learners exhibited significant differences between the white noise and babble noise conditions, with the F0 contour in the white noise being significantly higher than in the babble noise (all *ps* < 0.05). Furthermore, there were no significant differences between groups in either noise or quiet conditions.

For Tone 3, Bonferroni post-hoc tests showed that Korean participants made significant adjustments under both noise conditions compared with the quiet condition (all *ps* < 0.05). However, Chinese participants showed a significant difference only under the white noise condition compared with the quiet condition ($\beta = -0.08$, $SE = 0.03$, $t = -3.18$, $p = 0.02$), with no significant difference was found under the babble noise condition compared with the quiet condition. Both Chinese native speakers and Korean L2 learners exhibited significant differences between the white noise and babble noise conditions, with the F0 contour in the white noise being significantly higher than in the babble noise (all *ps* < 0.05). Furthermore, the results revealed that Korean L2 learners and Chinese speakers exhibited similar F0 contours under the quiet condition ($\beta = -0.19$, $SE = 0.07$, $t = -2.78$, $p = 0.08$), but their F0 contours differed significantly under the white noise condition ($\beta = -0.37$, $SE = 0.07$, $t = -5.37$, $p < 0.001$) and under the babble noise condition ($\beta = -0.30$, $SE = 0.07$, $t = -4.25$, $p < 0.001$). These findings suggested that noise conditions have different influences on native speakers and L2 learners. Figure 3 illustrates that under the two noise conditions, the overall F0 contours for Tone 3 were raised by Chinese speakers. However, for Korean speakers, the lowest point of the F0 contours increased more than other points, resulting in imprecise Tone 3 production.

For Tone 4, Bonferroni post-hoc tests showed that both Chinese and Korean participants had significant adjustments under the white noise conditions compared with the quiet condition (all *ps* < 0.05), while they showed similar results under the babble noise condition compared with the quiet condition. Furthermore, there were no significant differences between groups in either the noise or the quiet conditions.

Table 4: GAMMs Results for F0 Contours.

Term	Tone											
	Tone 1			Tone 2			Tone 3			Tone 4		
	<i>F</i>	<i>df</i>	<i>p</i>	<i>F</i>	<i>df</i>	<i>p</i>	<i>F</i>	<i>df</i>	<i>p</i>	<i>F</i>	<i>df</i>	<i>p</i>
Group	0.02	1	0.94	6.25	1	0.01	18.08	1	<0.001***	3.54	1	0.06
Noise	205.69	2	<0.001***	69.39	2	<0.001***	16.93	2	<0.001***	5.69	2	0.003**
Group: Noise	11.08	2	<0.001***	3.90	2	0.021*	13.17	2	<0.001***	16.74	2	<0.001***

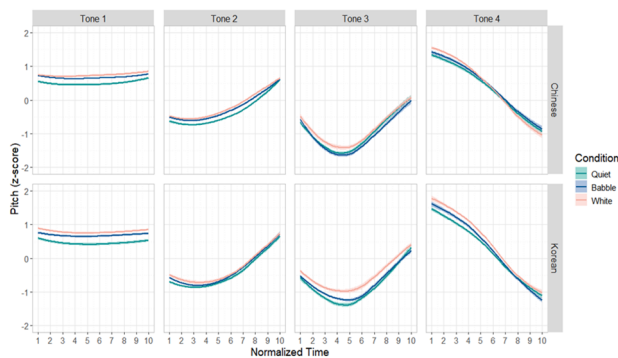


Figure 3: *F0 Contours for Speakers.*

4. Discussion

The results for intensity analysis showed that Chinese speakers did not exhibit any significant adjustments to the intensity of the four tones under the noise conditions compared with the quiet condition. These findings aligned with previous results, which indicated that Chinese speakers maintained consistent intensity under noise conditions [6]. This consistency might be attributed to the robustness of tonal information against noise [25], indicating that speakers did not need to significantly adjust their intensity in noisy environments to ensure accurate identification by listeners. In contrast, Korean L2 learners exhibited significant adjustments in the intensity of the four tones under noise compared to quiet conditions. These results were consistent with previous studies [7, 8]. This might be because L2 learners relied more heavily on auditory feedback control in speech production, resulting in more intensive monitoring of speech output and greater sensitivity to the speech environment compared with native speakers. As a result, L2 learners exhibited a greater magnitude of vocal compensation under noise conditions.

The duration results indicated that under the white noise condition, compared with the quiet condition, Chinese speakers only adjusted the speech duration of Tone 3 and Tone 4, demonstrating the Lombard effect. These results aligned with previous studies, which demonstrated that Chinese speakers had longer duration under noise conditions compared with quiet conditions [5, 6]. This might be because in Mandarin monosyllabic words, Tone 3 typically has the longest duration and Tone 4 has the shortest duration. In contrast, Korean L2 learners made significant adjustments for four tones under noise. This suggested a more pronounced Lombard effect among L2 learners than among native speakers, due to their greater dependence on feedback control in speech production.

The results for F0 contours indicated that both Chinese speakers and Korean L2 learners adjusted their F0 contours for all four tones under the noise conditions compared with the quiet condition. These findings were consistent with previous studies, which showed that both native speakers and Dutch L2 learners of English made pitch adjustments in the noise conditions compared with quiet conditions [10, 11]. Significant differences in adjustments between the two groups were observed only for Tone 3. While Korean L2 learners produced Tone 3 similarly to native speakers in quiet conditions, their speech production was significantly affected in the noise environment.

The results suggested that the white noise had a more significant effect than the babble noise. This could be due to two factors. First, white noise has a stronger masking effect

than babble noise due to a wider frequency range [26]. Secondly, both Chinese speakers and higher-level Korean L2 learners had more experience speaking in speech noise environments and less experience with white noise. Therefore, all participants were better able to maintain stable speech under babble noise compared with white noise.

These results have theoretical implications, contributing to our understanding of feedback control theory in L2 learners. Overall, the findings indicated that both native Chinese speakers and Korean L2 learners exhibited the Lombard effect, consistent with previous research demonstrating this effect in both L1 and L2 speakers [5, 6, 7]. Our study further revealed that the Lombard effect differed between native speakers and L2 learners. Native speakers relied more on the feedforward control system and less on monitoring their speech through the feedback control system. As a result, they were less affected by noise in terms of intensity and duration, and were able to maintain precise F0 contours, showing a smaller Lombard effect. The results for Korean L2 learners demonstrated two aspects of L2 speakers' reliance on feedback control. Firstly, the increased intensity and duration under noise suggested that feedback control in Korean learners prompted more adjustments when the auditory environment was altered, resulting in a greater Lombard effect than in native speakers. Secondly, the lower accuracy in F0 changes under the two noise conditions suggested that noise interfered with feedback control by disrupting the perception monitor of their own production. Under the quiet condition, the production of higher-level L2 listeners was similar with native speakers through a undisturbed feedback control. However, L2 listeners had lower accuracy and higher cognitive load in noise [27, 28], which affected their ability to monitor their own speech. In addition, although L2 learners were at a higher level, they had inadequate experience in noisy environments compared with native speakers, which might further limit their ability to process and adjust their speech accurately. The detrimental effect of noise on production was most evident for Tone 3, as it is considered the most articulatorily complex for non-tonal speakers [29]. This study suggested a compensatory relationship might exist among acoustic features, as Korean L2 learners relied more on intensity and duration due to difficulties in adjusting F0. Further research could explore the relationship between different acoustic features. Furthermore, this study focused on higher-level L2 learners; further research is needed on lexical tone production in lower-level L2 learners under noise.

5. Conclusions

This study investigates how native Mandarin speakers and Korean L2 learners adjust their production of lexical tones under different noise conditions, focusing on changes in intensity, duration, and F0 contours. Results showed that while native Mandarin speakers maintained consistent intensity across different conditions, Korean L2 learners significantly increased mean intensity in noise. In addition, Korean learners lengthened the duration for all tones, whereas native speakers only lengthened the duration of Tone 3 and Tone 4 under white noise. Differences between the two groups were noted in F0 contours for Tone 3. This study contributes to the understanding of feedback control theory in second language speech production.

6. References

- [1] M. Garnier and N. Henrich, "Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?" *Computer Speech and Language*, vol. 28, no. 2, pp. 580–597, 2014.
- [2] E. Lombard, "Le signe de l'elevation de la voix," *Ann Maladies de L'Oreille et du Larynx*, vol. 37, no. 2, pp. 101–119, 1911.
- [3] M. Vainio, D. Aalto, A. Suni, et al., "Effect of noise type and level on focus related fundamental frequency changes," in *Proc. INTERSPEECH 2012 – 13th Annual Conference of the International Speech Communication Association*, Portland, Oregon, USA, Sep. 2012.
- [4] K. Marcoux, M. Cooke, B. V. Tucker, and M. Ernestus, "The Lombard intelligibility benefit of native and non-native speech for native and non-native listeners," *Speech Communication*, vol. 136, pp. 53–62, 2022.
- [5] P. Mok, X. Li, J. Luo, and G. Li, "L1 and L2 phonetic reduction in quiet and noisy environments," in *Proc. Speech Prosody 2018 – 9th International Conference on Speech Prosody*, Poznań, Poland, Jun. 2018, pp. 1–5.
- [6] B. B. Li and J. Gong, "Comparison of Chinese and English rhythm produced by Chinese speakers in quiet and noise," in *Proc. IALP 2024 – 28th International Conference on Asian Language Processing*, Inner Mongolia, China, Aug. 2024, pp. 109–11.
- [7] X. Cai, Y. Yin, and Q. Zhang, "A cross-language study on feedforward and feedback control of voice intensity in Chinese–English bilinguals," *Applied Psycholinguistics*, vol. 41, pp. 771–795, 2020.
- [8] X. Cai, Y. Yin, and Q. Zhang, "Online control of voice intensity in late bilinguals' first and second language speech production: Evidence from unexpected and brief noise masking," *Journal of Speech, Language, and Hearing Research*, vol. 64, no. 5, pp. 1555–1570, 2021.
- [9] J. Villegas, J. Perkins, and I. Wilson, "Effects of task and language nativeness on the Lombard effect and on its onset and offset timing," *Journal of the Acoustical Society of America*, vol. 149, pp. 1855–1865, 2021.
- [10] K. Marcoux and M. Ernestus, "Differences between native and non-native Lombard speech in terms of pitch range," in *Proc. ICA 2019 and EAA Euroregio 2019– 23rd International Congress on Acoustics, Integrating 4th EAA Euroregio*, M. Ochmann, M. Vorländer, and J. Fels, Eds., Berlin, Germany, Sep. 2019, pp. 5713–5720.
- [11] K. Marcoux and M. Ernestus, "Pitch in native and non-native Lombard speech," in *Proc. ICPHS 2019 – 19th International Congress of Phonetic Sciences*, S. Calhoun et al., Eds., Melbourne, Australia, Aug. 2019, pp. 2605–2609.
- [12] J. Jin-xin, L. Yang, and Y. Xiao-hu, "Adjustment strategies in Mandarin tone production in noise among Korean learners," *Chinese Language Learning*, vol. 2022, no. 4, p. 92, 2022.
- [13] J. Kormos, "Individual differences in second language speech production," in *The Cambridge Handbook of Bilingual Processing*, J. W. Schwieter, Ed. Cambridge: Cambridge Univ. Press, 2015, pp. 369–388.
- [14] J. S. Perkell, "Movement goals and feedback and feedforward control mechanisms in speech production," *Journal of Neurolinguistics*, vol. 25, no. 5, pp. 382–407, 2012.
- [15] X. Li, Y. Huang, Y. Xu, L. Yi, Y. Yuan, and M. Xiang, "Production of Tone 3 Sandhi by advanced Korean learners of Mandarin," in *Proc. ISCSLP 2021 – 12th International Symposium on Chinese Spoken Language Processing*, Hong Kong SAR, China, Jan. 2021, pp. 1–5.
- [16] Y. Fu and Y. Lee, "The production of tone 3 by advanced Korean learners of Mandarin," *Linguistic Research*, vol. 39, no. 1, pp. 213–233, 2022.
- [17] Y. Xu, "ProsodyPro — a tool for large-scale systematic prosody analysis. Tools and resources for the analysis of speech prosody," in *Proc. TRASP 2013 – 1st Tools and Resources for the Analysis of Speech Prosody Conference*, Aix-en-Provence, France, Aug. 2013, pp. 7–10.
- [18] P. Rose, "Considerations in the normalisation of the fundamental frequency of linguistic tone," *Speech Communication*, vol. 6, pp. 343–351, 1987.
- [19] X.-N. Zhu, "F0 normalization: how to deal with between speaker tonal variations?" *Linguistic Science*, vol. 2, pp. 3–19, 2004.
- [20] K. Liu, L. Liang, and Q. Z. Qin, "The role of acoustic cues in the checked-unchecked tone merging of the Qixian Jin dialect," *Journal of the Acoustical Society of America Express Letters*, vol. 4, no. 12, p. 125202, 2024.
- [21] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
- [22] S. N. Wood, *Generalized Additive Models: An Introduction with R*, 2nd ed. Boca Raton, FL: Chapman and Hall/CRC, 2017.
- [23] R. V. Lenth, B. Banfai, B. Bolker, P. Buerkner, I. Giné-Vázquez, M. Herve, M. Jung, J. Love, F. Miguez, J. Piaskowski, H. Riebl, and H. Singmann, *emmeans: Estimated Marginal Means*, version 1.10.7, 2025.
- [24] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed., Springer, 2016.
- [25] H. Mixdorff, Y. Wang, and Y. Hu, "Robustness of tonal and segmental information in noise-auditory and visual contributions," in *Proc. Speech Prosody 2008 – 4th International Conference on Speech Prosody*, Campinas, Brazil, May 6–9, 2008, pp. 261–264.
- [26] M. L. Garcia Lecumberri and M. Cooke, "Effect of masker type on native and non-native consonant perception in noise," *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2445–2454, 2006.
- [27] A. Cutler, M. Cooke, M. L. Garcia Lecumberri, and D. Pasveer, "L2 consonant identification in noise: Cross-language comparisons," in *Proc. INTERSPEECH 2007 – 8th Annual Conference of the International Speech Communication Association*, Antwerp, Belgium, Aug. 27–31, 2007, pp. 1585–1588.
- [28] X. Yang and Y. Zhao, "L2 speech perception in noise," *Advances in Psychological Science*, vol. 22, no. 6, pp. 934–942, 2014.
- [29] L. R. Fabra, X. Liu, S. Chen, and R. Wayland, "Production of Mandarin tones by L1-Spanish early learners in a classroom setting," in *Second Language Speech Learning: Theoretical and Empirical Progress*, R. Wayland, Ed. Cambridge: Cambridge University Press, 2021, pp. 273–289.