A preliminary study of Mandarin neutral tone production by Japanese and Korean L2 learners

Tong Shu, Peggy Pik Ki Mok

The Chinese University of Hong Kong
tongshu@link.cuhk.edu.hk, peggymok@cuhk.edu.hk

ABSTRACT

This study investigates the production of Mandarin neutral tone by L2 learners whose L1s were Japanese and Korean. Mandarin neutral tone (T0) is characterized by its underspecified pitch contours and shorter duration compared with the four lexical tones (T1-T4). These two important acoustic correlates of Mandarin neutral tone, pitch and duration, are phonemic in Japanese (lexical pitch accent and vowel length contrast) while non-phonemic in Korean. Following the Feature Hypothesis, Japanese speakers were predicted to outperform Korean speakers in producing the two features of neutral tone. 10 Japanese and 10 Korean speakers with intermediate proficiency in L2 Mandarin participated in a reading task. The results showed that both groups had little difficulty producing the context-conditioned pitch patterns of neutral tone, whereas Japanese speakers produced more native-like duration patterns of neutral tone than Korean speakers did, which partially supported the Feature Hypothesis.

Keywords: L2 production, Mandarin neutral tone, Japanese, Korean, Feature Hypothesis

1. INTRODUCTION

The acquisition of L2 stress has been extensively studied. Stress is a suprasegmental feature realized by configuring several acoustic correlates such as F0, duration, and vowel quality. Previous studies on the phonetic implementation of L2 stress found that L2 learners have difficulties producing each acoustic correlate of L2 stress, which seems to be influenced by how these acoustic correlates are adopted in their native prosodic systems [1]–[3]. McAllister et al. (2002) [4] proposed the Feature Hypothesis concerning L1 transfer at the featural level. According to the Feature Hypothesis, if an L2 contrast is based on a phonetic feature which is not prominent in the L1, L2 learners will have difficulty perceiving and producing this contrast [4]. In their study, this is supported by the finding that Estonian speakers, whose L1 has phonemic vowel length contrast, produced L2 Swedish vowel length contrast in a native-like manner, compared with English and Spanish speakers whose L1s had limited or non-contrastive use of duration. Following the Feature Hypothesis, Lee and Guion (2006) [1] investigated the production of English unstressed vowels by early and late Japanese- and Korean-English bilinguals. Their study found that Japanese speakers produced comparable duration ratios of unstressed to stressed vowels with native English speakers, while Korean speakers did not. This result corresponded with the role of duration in their L1s: Japanese has phonemic vowel length contrast based on duration [5] whereas Korean has lost this contrast [6]. However, both Japanese and Korean groups were native-like in producing the F0 peak ratio of unstressed to stressed vowels, which did not echo the role of pitch in the two languages, i.e., phonemic in Japanese (lexical pitch accent) [7] but not in Korean [8].

In Mandarin, there is a 'toneless' category called the neutral tone. It resembles English unstressed syllables in that it has restricted pitch patterns, shorter duration, and vowel reduction [9], [10]. However, the acquisition of Mandarin neutral tone by L2 learners has received little attention [11], [12]. Therefore, the current study compared the production of L2 Mandarin neutral tone by L1-Japanese and L1-Korean speakers, whose L1s differed in the phonological status of duration and pitch, the two acoustic correlates of Mandarin neutral tone.

1.1. Types and acoustic correlates of neutral tone words

The neutral tone in Mandarin cannot occur independently and usually follows lexical tones. There are three typical types of disyllabic neutral tone words [13]: 1) the suffix type: some suffixes are usually produced in neutral tone, such as *-de* in *dui4de0* 'right', *-le* in *zou3le0* 'gone' and so on; 2) the reduplication type: words derived from reduplication usually have the second syllable in neutral tone, such as *ge1ge0* 'brother'; 3) the lexeme type: words that are conventionally pronounced with the second syllable in neutral tone, such as *zhi1shi0* 'knowledge'; Among the three types, neutral tone in the lexeme type is unpredictable and must be learned in a word-by-word manner, whereas the reduplication type and suffix type are predictable.

The acoustic correlates of neutral tone are F0, duration, and vowel quality [9], [10], [14]. Its pitch contour is determined by the preceding lexical tones,

which is falling after T1/2/4 but with different pitch heights and level/rising after T3 [15]. The duration of neutral tone is about 50~70% of the preceding tone, and some studies found that neutral tone after T3 tends to have larger duration ratios [13], [14]. Neutral tone showed varying degrees of vowel reduction compared with lexical tones, but the role of vowel reduction in native perception is less studied compared with F0 and duration. The acoustic properties of neutral tone reported in the above studies are summarized below.

Preceding lexical tone	Neutral tone	
	Pitch	Duration
T1 [55]	mid falling [41]	50~60%
T2 [35]	high falling [52]	50~60%
T3 [21]	mid level/rising [33/34]	~70%
T4 [53]	low falling [21]	50~60%

Table 1: Pitch patterns and duration of neutral tone.

Following the Feature Hypothesis and previous studies [1], Japanese speakers were predicted to 1) better distinguish the pitch contours of neutral tone following different lexical tones, and 2) produce more native-like duration patterns of neutral tone than Korean speakers would, given the more prominent role of pitch and duration in Japanese than in Korean.

2. METHOD

2.1. Participants

Ten L1-Japanese (NJ) and 10 L1-Korean (NK) L2 Mandarin learners, as well as 10 native Beijing Mandarin (NM) speakers participated in this experiment (see Table 2 for details). All the L2 learners were intermediate learners with equivalent proficiency of HSK-3~4 level, but an independent t-test indicated that the NK group had a significantly larger mean length of learning (LoL) than the NJ group, $t(16.2) = -2.65$, $p = .02 < .05$.

Group	Number	Age in yrs	LoL in yrs
NJ	6F, 4M	18.8 (1.03)	0.97 (1.69)
NK	7F, 3M	22.4 (3.41)	2.71 (1.19)
NM	7F, 3M	27.0 (7.60)	NA

Table 2: Summary of participant information. Inside brackets are standard deviations.

2.2. Materials

56 common disyllabic neutral tone words, including the three typical types introduced in Section 1.1, were selected as stimuli. All words are high-frequency words selected from common Chinese textbooks for L2 learners.

2.3. Procedures

Due to the COVID-19 pandemic, face-to-face experiments were difficult. Recent studies [16], [17] have suggested the feasibility of remote data collection via video conferencing software like ZOOM [18] and smartphone recordings. For duration and F0, the major concerns of this study, previous studies have shown that the two measurements remain stable across different devices [19]–[23]. Therefore, the current study combined ZOOM and smartphone recordings for data collection. The production stimuli were randomized and shown to the participants in PowerPoint Slides via the share-screen function in ZOOM. The participants were instructed to record themselves on their smartphones. All the stimuli were embedded in a carrier phrase *wo3du2 __ san1ci4* 'I read __ three times' to avoid final lengthening. Both simplified Chinese characters and pinyin were presented, and the participants were corrected when they mispronounced the preceding tones. Before the formal experiment, the participants were given careful instructions about the procedure, and they were required to record a demo for us to check and optimize the recording quality.

The preceding tone and the neutral tone in each token produced by the L2 learners were identified by two native Beijing Mandarin speakers. When the two judges give contradictory judgments, the first author would listen to the confusing token again and make the final judgment. Only tokens identified as carrying neutral tone were included in subsequent analysis, which made up 86.6% of the original dataset (2915 out of 3360 tokens).

2.4. Acoustic analysis

First, the onset and offset of the neutral tone and its preceding tone were marked manually in Praat [24]. The onset was identified as the beginning of the vowel, and the offset was identified as the end of the voiced interval. Second, the pitch and duration information were extracted using ProsodyPro [25]. For pitch analysis, the F0 values of neutral tone syllables at 10 evenly distributed time points were obtained and converted to z-scores for further visualization. For duration analysis, the absolute duration of the neutral tone and its preceding tone was converted into a duration ratio = neutral tone duration (ms) / preceding tone duration (ms) as in [13].

3. RESULTS

3.1. Pitch patterns of neutral tone

Figure 1 shows the pitch contour of neutral tone (T0 hereafter) following different lexical tones produced

by the NJ, NK, and NM groups visualized via the Smoothing Spline ANOVA model using the "gss" and "ggplot" packages in R [26], [27].

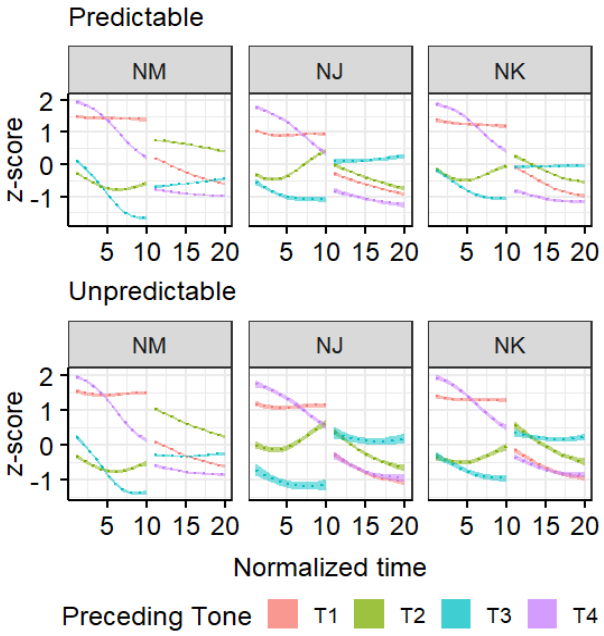

Figure 1: The pitch patterns of neutral tone after the four lexical tones produced by the NM, NJ and NK groups (Timepoint 1~10 represent the preceding lexical tone, and 11~20 represent the following neutral tone).

The pitch patterns of T0 produced by the NM speakers were consistent with previous studies (summarized in Table 1) for both the predictable and unpredictable types. Similarly, T0 produced by the NJ and the NK groups also clearly showed the falling vs. non-falling distinction between T0 after T1/2/4 vs. T3.

However, we can still observe several differences between the L2 learners and the NM speakers. First, the distribution of T0 after the four lexical tones in the NJ and the NK groups was generally less dispersed than in the NM group. Second, the relative position of T0 after T3 and the other T0s was different. In the NM group, T0 after T3 was placed in the middle position among all the T0s, while it had a higher overall pitch than the other T0s in both the NJ and the NK groups. Third, in terms of the relation between T0 and its preceding tone, in the NM group, the onset pitch of T0 after T2 was notably higher than the offset of its preceding tone, and the overall pitch level of T0 was higher than the preceding T2 as well. However, in both groups of L2 learners, the gap between the offset pitch of T2 and the onset pitch of the following T0 was much smaller than in the NM group.

Considering the effect of context, a difference in the shape of T0 after T3 can be observed between different contexts for both the NJ and the NK groups. It was level/rising in the predictable context while showing a slightly concave shape in the unpredictable context. Also, T0 after T1 and T4 largely overlapped in the unpredictable context, while they were better distinguished in the predictable context.

3.2. Duration ratios of neutral tone

A three-way mixed ANOVA was performed to examine the effects of Preceding Tone (T1 vs. T2 vs. T3 vs. T4), Type (predictable vs. unpredictable), and L1 (NJ vs. NK vs. NM) on the mean duration ratios of T0, with L1 as the between-subject factor, Preceding Tone and Type as the within-subject factors. The three-way interaction was significant, $F(6, 81) = 2.700$, $p = .02 < .05$. The simple two-way interaction between Group and Preceding Tone was significant for both the predictable type, $F(3.65, 49.29) = 5.954$, $p < .001$, and the unpredictable type, $F(6, 81) = 4.331$, $p < .001$. The simple one-way main effect of Group on mean duration ratios of neutral tone was statistically significant for the predictable type when the preceding tone was T2, $F(2, 27) = 11.960$, $p < .001$; T3, $F(2, 27) = 14.521$, $p < .001$; and T4, $F(2, 27) = 6.702$, $p = .004$. The simple one-way main effect of Group on mean duration ratios of neutral tone was statistically significant for the unpredictable type when the preceding tone was T2, $F(2, 27) = 6.673$, $p < .004$; T3, $F(2, 27) = 10.550$, $p < .001$; and T4, $F(2, 27) = 6.585$, $p = .005$.

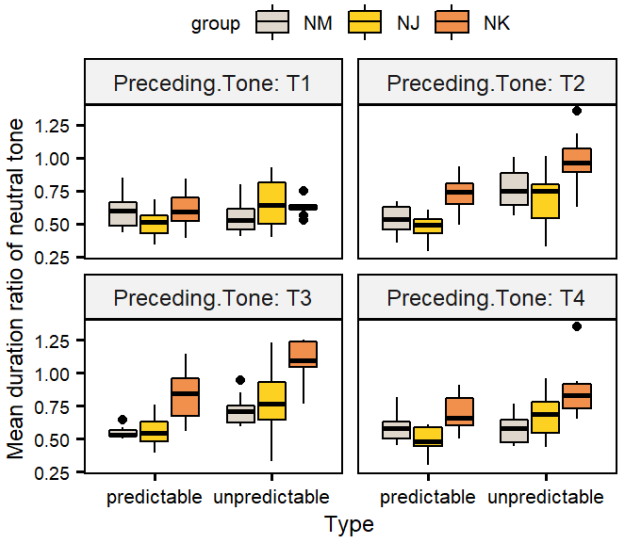

Figure 2: Mean duration ratios of neutral tone after four lexical tones produced by the three groups.

Pairwise comparisons (BH-adjusted) between groups were run where significant main effects were found. For T2 as the preceding tone, the NK group (*Mean* = 0.717, *SD* = 0.133) had significantly larger mean ratios than both the NM (*Mean* = 0.531, *SD* = 0.111) and the NJ group (*Mean* = 0.475, *SD* = 0.101), $p = .002 < .05$ and $p < .001$ respectively, while there was no significant difference between the NM and the NJ groups, $p = .29 > .05$. This result indicated that both the NJ and NM groups shorten the T0 syllable to

a similar degree while the NK speakers had longer T0 syllables. When the preceding tone was T3, the NK group (*Mean* = 0.841, *SD* = 0.202) had larger mean ratios than both the NM (*Mean* = 0.546, *SD* = 0.046) and the NJ group (*Mean* = 0.560, *SD* = 0.118), *p*s < .001, while the difference between the NM and the NJ groups was not significant, $p = .82 > .05$. When the preceding tone was T4, there was no significant difference between the NM (*Mean* = 0.587, *SD* = 0.117) and the NJ group (*Mean* = 0.491, *SD* = 0.102), $p = .09 > .05$, or between the NM and the NK groups (*Mean* = 0.689, *SD* = 0.140), $p = .09 > .05$. However, the NK group had significantly larger mean ratios than the NJ group, $p = .003 < .05$.

For the unpredictable type, when the preceding tone was T2, the NK group (*Mean* = 0.983, *SD* = 0.197) had larger mean ratios than both the NM (*Mean* = 0.762, *SD* = 0.162) and the NJ groups (*Mean* = 0.689, *SD* = 0.201), $p = .02 < .05$ and $p = .005 < .05$ respectively, while there was no significant difference between the NM and NJ groups, $p = .39 > .05$. When the preceding tone was T3, the NK group (*Mean* = 1.08, *SD* = 0.178) had larger mean ratios than both the NM (*Mean* = 0.715, *SD* = 0.112) and the NJ groups (*Mean* = 0.787, *SD* = 0.255), $p < .001$ and $p = .003 < .05$ respectively, while the difference between the NM and the NJ groups was not significant, $p = .41 > .05$. When the preceding tone was T4, the NK group (*Mean* = 0.859, *SD* = 0.203) had larger mean ratios than both the NM (*Mean* = 0.581, *SD* = 0.113) and the NJ groups (*Mean* = 0.681, *SD* = 0.191), $p = .004 < .05$ and $p = .045 < .05$ respectively, while the difference between the NM and the NJ groups was not significant, $p = .21 > 0.5$.

To conclude, for both types of T0 syllables, the NJ group's mean duration ratios of T0 were closer to the NM group than the NK group.

4. DISCUSSION

This preliminary study examined the production of L2 Mandarin neutral tone by Japanese and Korean speakers. In terms of pitch patterns, both the NJ and the NK speakers showed the context-conditioned pitch patterns of neutral tone like the NM speakers did, despite some target-deviant patterns observed in both groups. For example, the pitch patterns of neutral tone were consistent in the NM group regardless of context. In contrast, in both the NJ and the NK groups, the distribution of different neutral tones was less dispersed in the unpredictable context than in the predictable context. It was reasonable that neutral tone in the unpredictable context was more difficult, as it must be learned in a word-by-word manner, and was more likely to receive influence from its citation tone, compared with the other types, such as suffixes that do not have citation tones. Under the Feature Hypothesis [4], Japanese speakers were predicted to produce more accurate pitch contours of neutral tone than Korean speakers did because of the phonemic status of pitch in their L1. However, similar to Lee and Guion (2006)'s study [1], little difference was found between the NJ and the NK groups in terms of pitch. Note that there was a potential influence from the discrepancy in the length of learning between the two groups. The NK group had a longer mean length of learning than the NJ speakers did, which may obscure the cross-linguistic influence from native prosodic systems. However, as the NK group did not do better than the NJ group, the impact of the longer mean length of learning should be limited.

In terms of duration, the NJ speakers showed an advantage over the NK speakers in both the predictable and the unpredictable context. The NJ speakers' advantage over NK speakers in the predictable context was consistent with the Feature Hypothesis [4]. It is likely that NJ speakers' extensive experience manipulating duration to signify L1 vowel length contrast facilitated their production of the duration patterns of L2 Mandarin neutral tone. In addition, English learning experience may also facilitate the production of L2 Mandarin neutral tone. As mentioned in the Introduction, neutral tone resembles English unstressed syllables in terms of restricted F0, shorter duration, and vowel reduction. Therefore, they may adapt the duration pattern of English unstressed vowel to Mandarin neutral tone based on their similarity. It is possible since all the participants in the NJ and NK groups reported English learning experiences prior to Mandarin.

Only the hypothesis for duration but not pitch was supported in our study, which suggests that featural transfer based on phonemic status is not the whole story. More studies are needed to explore the potential factors modulating the transfer, such as the relative complexity of the prosodic contrast signified by the target feature in the L1 and L2. For example, in the current study, the L2 Mandarin tone has more extensive pitch variations than the L1 Japanese lexical pitch accent does. Thus, L1 experience with phonemic pitch may not necessarily facilitate the acquisition of a more complicated L2 contrast.

The present study has several limitations. First, we only compared the duration of neutral tone with its preceding tones, not with lexical tones occurring in the same positions, i.e., disyllabic-final. Second, we did not look at the influence of the onset type on the duration of neutral tone, which can be improved in future studies. Third, the sample size for each group was limited. To obtain a more comprehensive picture, more data from Japanese and Korean speakers with intermediate and higher proficiency is being collected.

5. ACKNOWLEDGEMENTS

The first author would like to thank Xiang Min, Zhao Ziyi, Zoe Cheung Sheung Yu for the help in data collection and annotation, Xu Huan for the guidance on statistical analysis, and all the participants who gave their time to participate in this study.

6. REFERENCES

[1] B. Lee, S. G. Guion, and T. Harada, "Acoustic analysis of the production of unstressed english vowels by early and late Korean and Japanese bilinguals," *Stud. Second Lang. Acquis.*, vol. 28, no. 3, pp. 487–513, 2006.

[2] W. Zuraiq and J. A. Sereno, "English Lexical Stress Cues in Native English and Non-Native Arabic Speakers," *ICPhS XVI*, no. August, pp. 829–832, 2007.

[3] Y. Zhang, S. L. Nissen, and A. L. Francis, "Acoustic characteristics of English lexical stress produced by native Mandarin speakers," *J. Acoust. Soc. Am.*, vol. 123, no. 6, pp. 4498–4513, Jun. 2008.

[4] R. McAllister, J. E. Flege, and T. Piske, "The influence of L1 on the acquisition of Swedish quantity by native speaker of Spanish, English and Estonian," *J. Phon.*, vol. 30, no. 2, pp. 229–258, 2002.

[5] Y. Hirata, "Effects of speaking rate on the vowel length distinction in Japanese," *J. Phon.*, vol. 32, no. 4, pp. 565–589, 2004.

[6] Y. Kang, T.-J. Yoon, and S. Han, "Frequency effects on the vowel length contrast merger in Seoul Korean," *Lab. Phonol.*, vol. 6, no. 3–4, pp. 469–503, 2015.

[7] S. Kawahara, "The phonology of Japanese accent," in *Handbook of Japanese phonetics and phonology*, M. Shibatani and T. Kageyama, Eds. De Gruyter, 2015, pp. 445–492.

[8] H.-S. Jeon, "Prosody," in *The handbook of Korean linguistics*, L. Brown and J. Yeon, Eds. John Wiley & Sons, Inc., 2015, pp. 41–58.

[9] M. Lin and J. Yan, "Beijinghua qingsheng de shengxue xingzhi[Acoustic properties of neutral tone in Beijing Mandarin]," *Fangyan*, no. 3, pp. 166–178, 1980.

[10] J. F. Cao, "Putonghua qingsheng yinjie texing fenxi[Analysis of the characteristics of Mandarin neutral tone syllables]," *Yingyong shengxue*, vol. 5, no. 4, pp. 1–6, 1985.

[11] C. B. Chang and Y. Yao, "Production of neutral tone in mandarin by heritage, native, and second language speakers," in *the 19th International Congress of Phonetic Sciences*, 2019, pp. 2291–2295.

[12] P. Tang, "Riben gaoji hanyu xuexizhe hanyu qingsheng yunlv xide pianwu fenxi[A study of prosodic errors of Chinese neutral tone by advanced Japanese students]," *Huawen jiaoxue yu yanjiu*, no. 4, pp. 39–47, 2014.

[13] P. Tang, I. Yuen, N. Xu Rattanasone, L. Gao, and K. Demuth, "Acquisition of weak syllables in tonal languages: Acoustic evidence from neutral tone in Mandarin Chinese," *J. Child Lang.*, vol. 46, no. 1, pp. 24–50, Jan. 2018.

[14] J. F. Cao, "On neutral-tone syllables in Mandarin Chinese," *Can. Acoust.*, vol. 20, no. 3, pp. 49–50, 1992.

[15] W.-S. Lee and E. Zee, "Prosodic characteristics of the neutral tone in Beijing Mandarin," *J. Chinese Linguist.*, vol. 36, no. 1, pp. 1–29, 2008.

[16] N. H. Hilton and A. Leemann, "Editorial: Using smartphones to collect linguistic data," *Linguist. Vanguard*, vol. 7, no. s1, pp. 1–7, 2021.

[17] A. Leemann, P. Jeszenszky, C. Steiner, M. Studerus, and J. Messerli, "Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing," *Linguist. Vanguard*, vol. 6, no. s3, 2020.

[18] I. Zoom Video Communications, "ZOOM cloud meetings." 2021, [Online]. Available: https://zoom.us/.

[19] C. Ge, Y. Xiong, and P. Mok, "How reliable are phonetic data collected remotely? Comparison of recording devices and environments on acoustic measurements," in *Proceedings of INTERSPEECH 2021*, 2021, pp. 3984–3988.

[20] P. De Decker and J. Nycz, "For the record: Which digital media can be used for sociophonetic analysis?," *Univ. Pennsylvania Work. Pap. Linguist.*, vol. 17, no. 2, p. Article 7, 2011.

[21] E. U. Grillo, J. N. Brosious, S. L. Sorrell, and S. Anand, "Influence of Smartphones and Software on Acoustic Voice Measures.," *Int. J. Telerehabilitation*, vol. 8, no. 2, pp. 9–14, 2016.

[22] C. Zhang, K. Jepson, G. Lohfink, and A. Arvaniti, "Comparing acoustic analyses of speech data collected remotely," *J. Acoust. Soc. Am.*, vol. 149, no. 6, pp. 3910–3916, 2021.

[23] Y. Guan and B. Li, "Usability and Practicality of Speech Recording by Mobile Phones for Phonetic Analysis," *2021 12th Int. Symp. Chinese Spok. Lang. Process. ISCSLP 2021*, 2021.

[24] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]." 2021, [Online]. Available: http://www.praat.org/.

[25] Y. Xu, "ProsodyPro - A tool for large-scale systematic prosody analysis," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody*, 2013, pp. 7–10.

[26] C. Gu, "Smoothing spline ANOVA models: R package gss," *J. Stat. Softw.*, vol. 58, pp. 1–25, 2014.

[27] H. Wickham, *ggplot2: elegant graphics for data analysis*. New York: Springer, 2009.