PRODUCTION OF NEUTRAL TONE IN MANDARIN BY HERITAGE, NATIVE, AND SECOND LANGUAGE SPEAKERS

Charles B. Chang & Yao Yao

Boston University, USA; The Hong Kong Polytechnic University, Hong Kong
cc@bu.edu; ctyaoyao@polyu.edu.hk

ABSTRACT

This study examined the properties of neutral tone (T0) in Mandarin as produced by three groups: native speakers raised in a Mandarin-speaking environment (L1ers), second language learners raised in an English-speaking environment (L2ers), and heritage language speakers (HLers) exposed to Mandarin from birth but currently dominant in English. T0 production was elicited in both obligatory and non-obligatory contexts, acoustically analyzed, and perceptually evaluated by Mandarin L1ers. Acoustic data indicated little difference among groups in pitch contour, but significant differences in duration, especially in the non-obligatory context. Perceptual data revealed relatively low intelligibility of T0 overall, but also a group difference whereby L2ers tended to outperform HLers in the non-obligatory context; nevertheless, L2ers received the lowest goodness ratings, across both contexts. These results thus suggest that phonetic differences between HLers and L2ers are not unidirectional, but instead vary across aspects of the language in accordance with differences in speakers' linguistic experience.

Keywords: heritage speakers, L2 learners, neutral tone, lexical tone, Mandarin Chinese.

1. INTRODUCTION

The burgeoning field of heritage language (HL) studies has led to a considerable amount of research on the phonetics and phonology of HLs. One theme in this literature is the phonetic and/or phonological ADVANTAGE that HL speakers (i.e., individuals who were exposed to the target language in childhood, but have since become dominant in a different language) tend to exhibit over late-onset second language (L2) learners [4, 14, 16, 18, 19]. However, another theme is that of DIVERGENCE between HL speakers and native (L1) speakers, whose acquisition of the language was not interrupted by early immersion in another language [1, 5, 6].

Although HL studies have examined a variety of phonetic features, prosody remains underexamined (cf. [13]), limiting the generalizability of previous conclusions about the patterning of HL speakers (HLers) vis-a-vis L1 and L2 speakers (L1ers, L2ers). For example, few studies have focused on HLers' knowledge and mastery of lexical tone, a feature of many HLs spoken around the world. Tone is of particular interest due to its early onset, but protracted development, in the L1 (cf. [24, 25, 26]). That is, it is a feature for which one might very well expect to see the type of "intermediate" patterning often observed for HLers relative to L1ers and L2ers.

In the present study, we build upon previous work on tone in HL, L1, and L2 varieties of Mandarin [5, 12, 27] to investigate the properties of Mandarin's neutral tone (T0), a short, "light" tone surfacing on weak syllables that "has no pitch value of its own, but acquires its pitch value according to context" [23]. Whereas Mandarin's four main tones (T1–T4) are the subject of extensive research, including work on dialectal variation [15, 17, 22], T0 has been less studied, with relatively little phonetic research even on L1 production (cf. [8, 11]). One reason for this may be the common analysis of T0 as the outcome of reduction (i.e., absence of T1–T4; [10, 28]). However, not all instances of T0 are amenable to such an analysis, as T0 does not necessarily alternate with one of T1–T4. That is, while some instances of T0 may be optional (e.g., *zi* in *ér zi* 'son', which may be produced with T0 or T3), others (especially in non-final position) are obligatory.

Thus, this study was aimed at making two contributions: (1) providing new phonetic data on T0 (crucially, from different contexts), and (2) drawing detailed, multi-measure comparisons among Mandarin speakers of different backgrounds. Our main question concerned whether the tonal advantages previously found for Mandarin HLers in the U.S., especially on T3 [5], would also be found on T0, a qualitatively different type of tone. We hypothesized that, due to the fact that formal Chinese classes typically teach T0 explicitly but tend not to distinguish between obligatory and non-obligatory T0 contexts, Mandarin HLers and L2ers would diverge in their T0 production in non-obligatory contexts. In particular, L2ers taught Standard Mandarin (largely based

on northern varieties, which tend to realize T0 in non-obligatory contexts) were predicted to produce T0 consistently in non-obligatory contexts, whereas HLers (with more exposure to southern varieties and less educational exposure to Standard Mandarin) were predicted to show more variability.

To test these predictions, we conducted a small-scale study of T0 production by Mandarin HLers, L2ers, and L1ers, collecting two types of data. Acoustic data on fundamental frequency (f_0) contour and duration were gathered to address our central predictions regarding consistency of T0 production, while perceptual data were gathered to further evaluate the quality of the productions.

2. METHODS

2.1. Participants

Three groups of Mandarin speakers in the U.S. (California) participated in the production study: native Mandarin speakers (L1ers), late-onset L2 learners (L2ers), and HL speakers (HLers) representing a range of exposure to Mandarin. L1ers ($N = 6$; 4f, 2m; $M_{age} = 29.8$ yr, $SD = 8.5$) were born and educated in Mainland China (4) or Taiwan (2) until at least seventh grade and were late arrivals to the U.S. ($M_{AOA} = 24.2$ yr, $SD = 8.1$). In contrast, L2ers ($N = 5$; 3f, 2m; $M_{age} = 21.6$ yr, $SD = 3.7$) were born and educated in the U.S. and raised in English-speaking families; they had started to learn Mandarin after age 18 (through instruction and/or prior travel to a Mandarin-speaking country) and generally described their proficiency at the time of testing as relatively poor (e.g., self-assessments of conversational comprehension ranging from 10% to 50%).

HLers ($N = 15$) were born to Mandarin-speaking parents, but reported speaking English most of the time overall and did not meet the description of L1ers (i.e., being raised in a Mandarin-speaking country until adolescence, perceiving their Mandarin proficiency to be native-like, and identifying as dominant in Mandarin). Given the wide range in Mandarin exposure and use, HLers were further divided into two subgroups for analysis. The high-exposure (HE) group ($N = 9$; 4f, 5m; $M_{age} = 21.0$ yr, $SD = 1.7$) reported using Mandarin to communicate with both parents most or all of the time, whereas the low-exposure (LE) group ($N = 6$; 4f, 2m; $M_{age} = 20.0$ yr, $SD = 1.1$) reported using Mandarin at home half of the time or less and had mostly never lived in a Mandarin-speaking country.

The listeners who served as judges for perceptual rating were L1 Mandarin speakers born, raised, and educated primarily in Mainland China ($N = 64$; 47f, 17m; $M_{age} = 23.7$ yr, $SD = 4.2$) who were living in Hong Kong at the time of testing.

2.2. Materials

The materials for the production study comprised 59 items, of which 6 were critical items targeting T0, 16 were items targeting T1–T4, and 37 were fillers and items included as part of other studies not discussed here. The critical items contained common words likely to be familiar to the participants and, crucially, included T0 in both obligatory contexts (i.e., in morphemes that must be produced with T0) and non-obligatory contexts. The obligatory T0 items were (in pinyin; morphemes with T0 underlined): *hē le shuǐ* 'drink water' + ASPECT, *chī le fàn* 'eat food' + ASPECT, *nǐ de shū* 'your book', and *hǎo kàn de rén* 'good-looking person'. The non-obligatory T0 items were: *zhuō zi* 'table' and *ér zi* 'son'.

The stimuli submitted to perceptual rating consisted of the speech recorded in the production study. The set of critical stimuli thus comprised 624 sound files (26 talkers × 6 items × 4 tokens), with one excluded from analysis due to file corruption.

2.3. Procedure

This study consisted of a Mandarin production task and a rating task. Talkers in the production study first completed a background questionnaire (adapted from [9]) and then a reading task in a sound booth. In this task, items were presented in random order on flashcards, which included Chinese characters (simplified or traditional) and romanization (pinyin and/or Zhuyin/Bopomofo). Talkers were told to produce the items naturally, and were audio-recorded at 48 kHz and 16 bps, using an AKG head-mounted condenser microphone connected either to a Marantz PMD660 or to a Dell desktop computer (through an M-AUDIO USB preamp).

The rating task, administered via ExperimentMFC in Praat [3] on a Lenovo ThinkPad X240 laptop using headphones, consisted of two parts, with five response options for both: T0–T4 for tone identification, and 1–5 (5 = high) for tonal goodness rating. Because the stimuli were evaluated syllable by syllable, they were organized into blocks according to syllable count. On each trial, listeners identified the tone in the current syllable of the given stimulus (played in its entirety), and then rated the goodness of that tone. Due to the number of stimuli, they were distributed among four versions of the task (with listeners completing only one version), such that each stimulus was evaluated by a panel of about eight different listeners.

2.4. Analysis

Recordings underwent acoustic analysis in Praat as in [5]. First, the recordings were annotated for the onset and offset of the voiced interval over which an f_0 contour would be extracted; this was done via auditory inspection and joint visual inspection of the waveform and a wide-band spectrogram (on the basis of criteria such as changes in periodicity, amplitude, and formant structure), according to the segments in the item. Measurements of voiced interval durations and of f_0 at 10 evenly spaced time points (ranging from the 5% to the 95% point) were then extracted via cross-correlation (default settings used except that the voicing threshold was set to 0.25 and the pitch floor was adjusted by talker to provide the best f_0 tracking possible).[1] All f_0 measurements were then log-transformed and converted to the T scale [21, 29] ranging from 0 to 5 (cf. the five-point representation system of [7]).[2]

The statistical analysis of both acoustic and perceptual data was done with mixed-effects modeling in R [20], using *lme4* [2]. All models had the same basic structure, including random effects for Talker and Item and fixed effects for Group (L1er, HLer-HE, HLer-LE, L2er; baseline = L2er), Context (obligatory, non-obligatory; baseline = obligatory), and their interaction. In this study, the critical effect in each model is the Group × Context interaction, since our primary concern is whether the various groups differ in T0 production across contexts.

3. RESULTS

3.1. Acoustic data: f_0 contour and duration

Figure 1 plots the average f_0 contour of T0 for all talkers, labeled by group. As shown, there is a high degree of similarity across groups in terms of the overall shape of the T0 contour. To be specific, most talkers produced a falling contour, which tended to be relatively shallow (compared to the steep fall characteristic of other tones such as T4) although some L2ers and HLers showed a steeper fall.

As expected, the duration of T0 (i.e., of the voiced interval in target syllables with T0), averaging 133 ms across groups, was considerably shorter than that found for T1–T4 (cf. mean durations of 191–295 ms in multisyllabic items in [5]). Nevertheless, T0 duration varied across groups and contexts, as shown in Figure 2.[3] Raw T0 durations were log-transformed and then analyzed in a linear mixed model. This model showed no significant between-group differences in the obligatory context ($|t| < 2$). However, in the non-obligatory context, whereas there was no

Figure 1: Average T0 contours, by group.

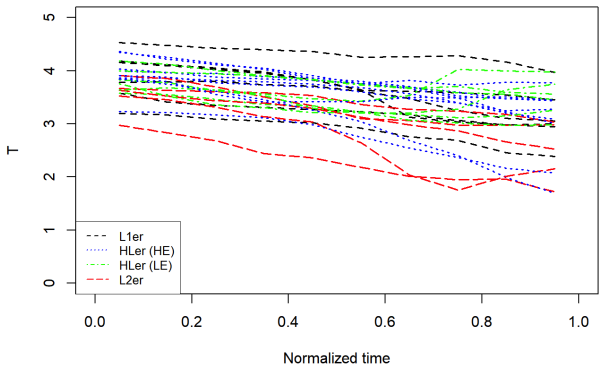
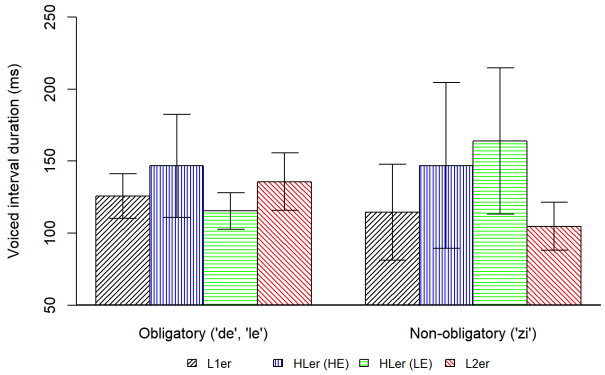


Figure 2: T0 duration, by context and group.



significant difference between L1ers and L2ers, both HLer groups produced significantly longer durations than L2ers ($\beta_{HE:nonobl} = 0.222, t = 3.388$, 95% $CI = \{0.094, 0.350\}$; $\beta_{LE:nonobl} = 0.559$, $t = 7.872$, 95% $CI = \{0.420, 0.699\}$). Thus, L2ers, but not HLers, resembled L1ers in terms of producing relatively short T0 durations in the non-obligatory context.

3.2. Perceptual data: intelligibility and goodness

While the global intelligibility of T0 (i.e., likelihood of accurate identification by L1 listeners) averaged over all talkers was about 60%, and thus lower than that reported for T1–T4 (cf. 78–92% in [5]), T0 intelligibility showed variation across contexts and groups (Figure 3). A logistic mixed model showed that in the obligatory context both the L1er group ($\beta = 0.802, Z = 2.246, P = .025$) and the LE (HLer) group ($\beta = 0.829, Z = 2.322, P = .020$) produced T0 more intelligibly than L2ers, while the difference between the HE (HLer) and L2 groups did not reach significance ($\beta = 0.421, Z = 1.286, P = .199$). However, in the non-obligatory context, the pattern was

Figure 3: T0 intelligibility, by context and group.

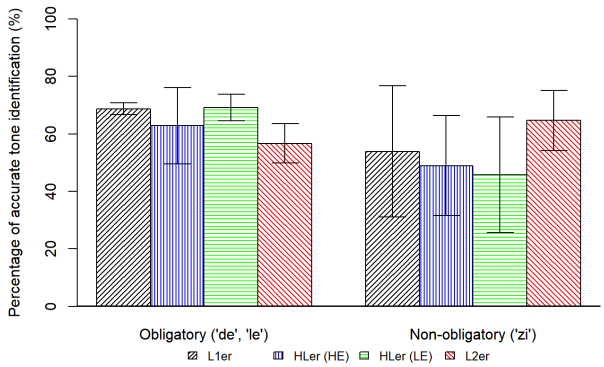
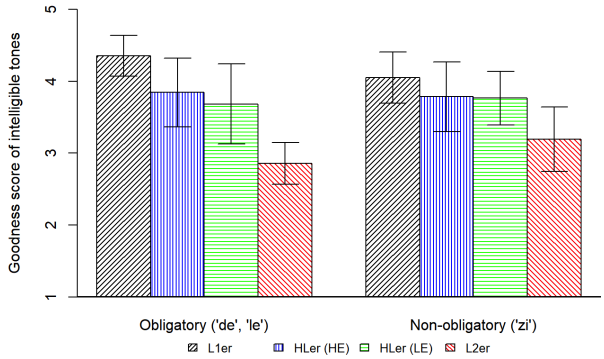


Figure 4: T0 goodness, by context and group.



reversed ($\beta_{L1er:nonobl} = -1.471, Z = -6.236, P < .001$; $\beta_{HE:nonobl} = -1.386, Z = -6.397, P < .001$; $\beta_{LE:nonobl} = -1.977, Z = -8.373, P < .001$), with L1ers and HLers both showing lower T0 intelligibility than L2ers.

As for the perceived goodness of T0 (i.e., the quality ratings listeners gave to T0 productions that they identified correctly as T0), this measure also showed variation across groups, within the same range of scores given to T1–T4 productions in [5]. As depicted in Figure 4, group variation showed a similar pattern across the two contexts. Results of a linear mixed model showed that, in the obligatory context, L1ers and HLers both received higher goodness ratings than L2ers ($\beta_{L1er} = 1.493, t = 5.875$; $\beta_{HE} = 0.992, t = 4.230$; $\beta_{LE} = 0.842, t = 3.315$). In the non-obligatory context, the between-group differences vis-a-vis L2ers were reduced ($\beta_{L1er:nonobl} = -0.606, t = -5.172$; $\beta_{HE:nonobl} = -0.444, t = -4.039$; $\beta_{LE:nonobl} = -0.246, t = -2.038$), but not enough to change the directionality of any difference; thus, here, too, it was the case that L1ers and HLers received higher ratings than L2ers.

4. GENERAL DISCUSSION

Taken together, these results paint a picture in which Mandarin HLers do not consistently have the "upper hand" over L2ers in terms of patterning like L1ers; rather, the manner in which HLers and L2ers differ depends on an interaction between the aspect of the language at issue (e.g., non-obligatory T0 contexts) and convergence in relevant experience and/or knowledge with L1ers (e.g., exposure to and familiarity with standard norms). Thus, whereas few systematic differences were observed among groups in obligatory T0 contexts, there were several differences observed in non-obligatory contexts, because it is in these contexts where there are relevant experiential disparities among the groups. Here, L2ers (presumably due to greater exposure to a variety in which T0 is produced at high rates even in non-obligatory contexts) showed evidence of more consistent T0 production than HLers, which therefore led to an advantage over HLers in T0 intelligibility. Interestingly, however, HLers still maintained an advantage in T0 goodness; that is, when they did manage to produce T0 intelligibly, HLers' production was perceived as higher-quality than L2ers'.

As the product primarily of differences in dialectal experience, the current results raise a number of questions for further research related to language variation and change in HLs. For one, although our interpretation of the observed HLer-L2er disparities is consistent with the fact that (according to background questionnaires) more than half of the HLers did indeed have substantial exposure to southern varieties of Mandarin, it remains unclear what role variation in dialectal experience (in particular, northern vs. southern in this case) might play in accounting for other group disparities (e.g., the L1er advantage in perceived goodness). In addition, the composition of our listener pool, which consisted exclusively of L1ers, naturally invites the question of what perceptual judgments from another group, such as HLers, would look like in comparison (e.g., how is northern/southern dialectal variation in Mandarin perceived by HLers in comparison to how such variation is perceived by homeland L1ers?).

In conclusion, our findings highlight the importance of two variables in the study of HL phonetics: CONTEXT and EXPERIENCE. This is not the first, and is unlikely to be the last, study of HLers to show a context effect, as well as a dialect/education effect, on the patterning of between-group differences. The challenge for future research will be to carefully tease these effects apart in order to catch a glimpse of HL behavior that truly reflects HL knowledge.

5. REFERENCES

[1] Ahn, S., Chang, C. B., DeKeyser, R., Lee-Ellis, S. 2017. Age effects in first language attrition: Speech perception by Korean-English bilinguals. *Language Learning* 67(3), 694–733.

[2] Bates, D., Maechler, M., Bolker, B. 2011. Lme4: Linear mixed-effects models using S4 classes [R package]. Version 0.999375-39. Available from http://cran.r-project.org/package=lme4.

[3] Boersma, P., Weenink, D. 2016. Praat: Doing phonetics by computer. Version 6.0.19. http://www.praat.org.

[4] Chang, C. B. 2016. Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition* 19(4), 791–809.

[5] Chang, C. B., Yao, Y. 2016. Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin. *Heritage Language Journal* 13(2), 134–160.

[6] Chang, C. B., Yao, Y., Haynes, E. F., Rhodes, R. 2011. Production of phonetic and phonological contrast by heritage speakers of Mandarin. *JASA* 129(6), 3964–3980.

[7] Chao, Y. R. 1930. A system of "tone-letters". *Le Maître Phonétique* 45, 24–27.

[8] Chen, Y., Xu, Y. 2006. Production of weak elements in speech – Evidence from F_0 patterns of neutral tone in Standard Chinese. *Phonetica* 63(1), 47–75.

[9] Dai, J. E., Zhang, L. 2008. What are the CHL learners inheriting? *Habitus* of the CHL learners. In: He, A. W., Xiao, Y. (eds.), *Chinese as a Heritage Language: Fostering Rooted World Citizenry*. Honolulu: NFLRC, 37–51.

[10] Duanmu, S. 2007. *The Phonology of Standard Chinese* (2nd ed.). Oxford: Oxford University Press.

[11] Fan, S., Li, A., Chen, A. 2018. Perception of lexical neutral tone among adults and infants. *Frontiers in Psychology* 9, 322.

[12] Hao, Y.-C. 2012. Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *JPhon* 40(2), 269–279.

[13] Kim, J. Y. in press. Discrepancy between heritage speakers' use of suprasegmental cues in the perception and production of Spanish lexical stress. *Bilingualism: Language and Cognition*.

[14] Knightly, L. M., Jun, S.-A., Oh, J. S., Au, T. K. 2003. Production benefits of childhood overhearing. *JASA* 114(1), 465–474.

[15] Lee, L. 2010. The tonal system of Singapore Mandarin. In: Clemens, L. E., Liu, C.-M. L. (eds.), *Proceedings of NACCL-22 and IACL-18*, vol. 1. Cambridge: Harvard University, 345–362.

[16] Lee-Ellis, S. 2012. *Looking into Bilingualism through the Heritage Speaker's Mind*. PhD thesis, University of Maryland, College Park, MD.

[17] Lin, Y. 2018. *Stylistic Variation and Social Perception in Second Dialect Acquisition*. PhD thesis, Ohio State University, Columbus, OH.

[18] Lukyanchenko, A., Gor, K. 2011. Perceptual correlates of phonological representations in heritage speakers and L2 learners. In: Danis, N., Mesh, K., Sung, H. (eds.), *Proceedings of BUCLD 35*, vol. 2. Somerville: Cascadilla Press, 414–426.

[19] Oh, J., Jun, S.-A., Knightly, L., Au, T. 2003. Holding on to childhood language memory. *Cognition* 86(3), B53–B64.

[20] R Development Core Team, 2016. R: A language and environment for statistical computing. Version 3.3.3. http://www.r-project.org.

[21] Shi, F. 1986. Tianjin fangyan shuangzizu shengdiao fenxi [An analysis of the bisyllabic tones in Tianjin dialect]. *Yuyan Yanjiu* 1986.1, 77–90.

[22] Torgerson, R. C. 2005. A comparison of Beijing and Taiwan Mandarin tone register: An acoustic analysis of three native speech styles. Master's thesis, Brigham Young University, Provo, UT.

[23] Wang, J. 2004. The neutral tone in trisyllabic sequences in Chinese dialects. In: *Proceedings of TAL-2004*. Beijing, China: International Speech Communication Association 201–202.

[24] Wong, P. 2013. Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *JASA* 133(1), 434–443.

[25] Wong, P., Schwartz, R. G., Jenkins, J. J. 2005. Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *JSLHR* 48(5), 1065–1079.

[26] Wong, P., Strange, W. 2017. Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study. *PLoS ONE* 12(8), e0182337.

[27] Yang, B. 2015. *Perception and Production of Mandarin Tones by Native Speakers and L2 Learners*. Berlin: Springer Verlag.

[28] Zhang, J. 2007. A directional asymmetry in Chinese tone sandhi systems. *JEAL* 16(4), 259–302.

[29] Zhu, X. 2004. Jipin guiyihua – ruhe chuli shengdiao de suiji chayi [F0 normalization: How to deal with between-speaker tonal variations?]. *Yuyan Kexue* 2004.3(2), 3–19.

[1] All f_0 measurements were further inspected for tracking errors, and obvious errors (which occurred in 23% of tokens) were hand-corrected. To correct contours containing large pitch jumps and/or gaps, the cross-correlation settings were adjusted; however, when a contour resisted correction via adjustment of analysis settings, an f_0 measurement was calculated manually by taking the duration over a 2–3 period interval around the relevant time point and converting to an f_0 value.

[2] All data and materials associated with this study are publicly accessible via the Open Science Framework at https://osf.io/u4w2g/.

[3] Here and elsewhere, error bars represent ± 1 *SE* of the mean over participants. Furthermore, because the two morphemes representing the obligatory context always showed the same pattern, they have been plotted together.