L2-Mandarin regional accent variability during lexical tone word training facilitates naive English listeners' tone categorization and discrimination

Yanping Li[1], *Catherine T. Best*[1,2], *Michael D. Tyler*[1,3], *Denis Burnham*[1]

[1] The MARCS Institute, Western Sydney University, Australia
[2] Haskins Laboratories, New Haven, U.S.A.
[3] School of Psychology, Western Sydney University, Australia

yanping.li@westernsydney.edu.au, c.best@westernsydney.edu.au
m.tyler@westernsydney.edu.au, denis.burnham@westernsydney.edu.au

Abstract

Mandarin-naïve English listeners have difficulties categorizing the four lexical tones that distinguish word meanings in Mandarin. This study investigated how L2-Mandarin regional accent variability in training on minimal-tone-contrast words affected tone perception. Prior to training, although listeners accurately categorized and discriminated rising and dipping tones, they confused falling and level tone significantly more than the other tone contrasts. After training, learners in the accent variability (experimental) condition showed improved categorization and discrimination of falling and level tones; constant-accent (control condition) learners did not. The results supported the hypothesis that accent variability during lexical tone word training facilitates tone categorization.

Index Terms: high vs low variability training; regional accents; lexical tone discrimination; tone contour categorization

1. Introduction

There are four lexical tones in Mandarin [1] that differ in their fundamental frequency (f_0) patterns with f_0 height and f_0 contour as the primary acoustic parameters[2]: T1 high level contour, T2 high rising, T3 low dipping and T4 high falling. Tones are used to distinguish word meanings (e.g., for the consonant-vowel [CV] syllable /ma/, level = *mother*, rising = *hemp*, dipping = *horse*, and falling = *to scold*). Listeners of languages that lack tones at the pre-lexical level, e.g., English or French listeners, perceptually assimilate tones to their native intonational categories [3], [4]. For example, the rising and falling tones are assimilated to English question (yes-no) versus statement intonations, respectively, due to phonetic similarities. Thus, non-tone language listeners are not entirely "deaf" to tones. When categorizing tones according to perceived pitch/contour, without referring to native intonational categories, their tone perception is psychophysically based [5], reflected in varying performance across tones [6]–[8]. Specifically, while listeners can categorize the pitch/contour of tone contrasts that display acoustic dissimilarities, such as T1 level vs. T3 dipping, T2 rising vs. T4 falling and T3 dipping vs. T4 falling, they have difficulties with those that show acoustic similarities, e.g., T1 level vs. T4 falling, T1 level vs. T2 rising, and T2 rising vs. T3 dipping [9].

Despite initial difficulties in categorizing Mandarin tones, naive English listeners can show improvements after high-variability perceptual training with multiple talkers. In [10], tone categorization improved 21% from pre- to post-test after high-variability training, which was maintained 6 months later. They also showed generalization to novel stimuli from one of the talkers used in training and to a novel talker. Based on this result, high-variability perceptual training has been adapted in other second language (L2) suprasegmental training studies (e.g., [11], [12]).

One limitation of perceptual training for improving L2 acquisition is that the training focuses on tone categorization rather than their lexical relevance, i.e., tone word identification. Meaningful words, rather than isolated (supra)segments, are employed in conversations, which creates coordinated patterns of activity in sensory and higher level cognitive functions [13]. In tasks involving simple tone categorization, even high-variability perceptual training still fails to relate tonal form to lexical meaning. Training with words that contrast only in their lexical tones can address this shortcoming. In a picture-to-word L2 training paradigm, English learners trained on Spanish words produced by multiple talkers showed significantly better accuracy in identifying target words than those trained with just a single talker [14]. Thus, high talker variability can facilitate L2 word learning in a non-tone language. One study of potential relevance to talker variability effects on minimal-tone-contrast word learning trained English learners to identify pseudowords recorded by English speakers, in which the pitch contours had been resynthesized. Participants with high-variability (multiple talkers) training achieved higher post-training accuracy than those with low-variability (one talker) training [15].

If high talker variability word training helps English listeners identify minimal-contrast words differing only in tone, it should in turn promote tone categorization and discrimination for each of the six tone contrasts (T1 vs. T2, T1 vs. T3, T1 vs. T4, T2 vs. T3, T2 vs. T4, T3 vs. T4), which was not investigated in prior studies either on tone perception [6]–[9] or on Mandarin minimal-tone-contrast word training [15], [16]. If English learners can identify words differing only in tone after training, implying that they have established tone categories, they should be able to use those categories to sort and discriminate tones at the pre-lexical level. We tested this hypothesis by examining tone categorization and discrimination after word training.

While our training used high talker variability, we manipulated degree of accent variability, which has not been examined before. Accent variability in tones is triggered by similarities and dissimilarities between the tone systems of Chinese regional dialects and standard Beijing Mandarin [17], [18] which regional speakers learn as an L2. For example, Yantai, Shanghai and Guangzhou speakers produce their L2 Mandarin dipping tone with shallower falling-rising contours than Beijing Mandarin productions (see [18] for acoustic details). This study investigates how L2-Mandarin regional

accent variability affects English listeners' L2 tone categorization and discrimination following word training. We posit that exposure to L2-Mandarin regional accent variability will facilitate word learning.

In our study, English participants in two conditions were trained on minimal-tone-contrast words with high talker variability. The control group heard only Beijing-accented stimuli, whereas the experimental group heard them spoken with Beijing and another two L2-Mandarin regional accents. Their tone contour categorization and discrimination in pre- and post-training conditions were estimated under the framework of Perceptual Assimilation Model (PAM, [19], [20]), which focuses on perceptual assimilations to native phoneme categories in cross-language speech perception by (naïve) L2 listeners and PAM has been extended to lexical tone assimilation [21], [22]. Prior to training, English listeners should perceptually assimilate tones as Non-Assimilable nonspeech pitch heights and contours, because there are no lexical tone categories in their phonological system. Nonspeech tone icons presented in [22] will be used for categorization. Naïve English listeners should assimilate each tone to several tone icons sharing height/contour similarities, which are learnable according to PAM-L2 [19]. Given this and four icons, although they are expected to be initially Non-Assimilable to phonological categories, they can nonetheless be either Categorized or Uncategorized to specific nonspeech icons. Their discrimination of the four Mandarin tones is expected to be good to very good. After minimal-tone-contrast word training, learners in both conditions should be able to abstract the tones as lexically relevant L2 phonological categories and their assimilation types are expected to shift to being Categorized or Uncategorized as L2 phonological components in speech with experimental listeners performing better than the control group due to effects of L2-Mandarin regional accent variability. Their tone discrimination varies based on tone assimilation types and training conditions. Specifically, listeners in the experimental group are more likely to assimilate each tone to a different tone contour category and their discrimination will be excellent. Listeners in the control group should easily Categorize tones with dissimilarities. On the contrary, they may show Uncategorized assimilation of tones with height/contour similarities, such as T1 level vs. T4 falling, resulting in Categorized-Uncategorized assimilation with good discrimination.

2. Experiment

2.1. Method

2.1.1. Participants

Mandarin-naïve English speakers (n = 48) were recruited online for this study, and randomly assigned to the single accent (control: n = 24, M_{age} = 24.5 years, SD = 5.8 years, 14 females) or multiple accent (experimental: n = 24, M_{age} = 25.5 years, SD = 5.1 years, 15 females) conditions. All were functional monolinguals [9] from English-speaking countries, primarily Australia. Prior to this study, none had experience with any tone languages, e.g., Mandarin, Vietnamese, Thai, Cantonese. Since musical training can facilitate tone perception [15], [23], none had more than 3 years of private lessons in any combination of instruments [15]. Their language and music backgrounds were self-reported through an online survey, which also determined that none had speech or hearing disorders. They received Prezzee eGift smart cards for their participation.

2.1.2. Stimuli

There were 16 Mandarin tone real words for pre- vs. post-categorization and discrimination tasks, which were generated from four CV syllables (/ga/, /ti/, /tu/, /pu/) with four tones. The three vowels /a/, /i/, /u/ were selected, because they are used in both Mandarin and English. Word targets and their characters were very regular based on [24] and they were produced by a native Beijing female talker (Age = 25.0 years) in a soundproof booth at the Speech Acquisition and Intelligent Technology Lab, Beijing Language and Culture University, Beijing, China. Apparatus and procedures for recordings were the same as those in [18]. Her productions were verified by four other native Beijing female listeners (M_{age} = 20.75 years, SD = 2.49 years) on a scale of 1 (not clear) to 7 (clear) and only the four tokens with the highest ratings for each word were retained, resulting in 64 (16 words × 4 tokens) stimulus items.

Another set of 16 Mandarin tone real words (four CV syllables, /ba/, /di/, /du/, /gu/ × 4 tones) were used in the Mandarin tone word training task. They were produced by native female talkers selected from [18], either 12 from Beijing (control: constant accent, talker-only variability) or four each from Beijing, Yantai, and Guangzhou (experimental: accent and talker variability). Both groups thus heard 12 talkers. As with the pre- and post-training stimuli, training word productions were verified by four other native female listeners of each dialect and final tokens were selected, resulting in 768 (12 speakers × 16 words × 4 tokens) stimuli in each training group. Word meanings for training were indicated by grey-scale pictures from [25], counterbalanced across participants, resulting in 16 pseudowords per participant.

2.1.3. Procedure

This study was conducted remotely with E-prime Go 1.0. Participants ran the perceptual tasks on their own Windows 10 laptops/desktops. To ensure data quality, they were tested through a ZOOM meeting with the experimenter, using wired (not bluetooth/wireless) earphones in a quiet room.

Learners completed Mandarin tone word training, pre- and post-training tone categorization and discrimination tasks, and post-training word verification and generalization tests (not reported here). The focus of this study was how L2-Mandarin regional accent variability affects the listeners' pre- to post-training tone perception, so the training procedure is described here briefly. They learned the 16 tone pseudowords in a picture-to-word paradigm, which were produced by 4 talkers in each training session (45 minutes), with the selected groups of talkers counterbalanced across the six training sessions that used quizzes with feedback [15]. Experimental group talkers were blocked by regional accent. Correspondingly, talkers for the control group were randomly assigned to subgroups of four with the *R* [26] *Sample* function.

A 1-minute tone familiarization was presented at the start of the pre-training perceptual test to acquaint listeners with the four tone contour icons. They then completed the tone discrimination and categorization tests in that order. As in studies investigating PAM-based predictions for consonants [27] and vowels [28], in the pre- and post-training perceptual tests, listeners first completed a categorical AXB discrimination test for each of the six tone contrasts, which were blocked with a Latin Square design across participants. In each trial, A and B were tokens of the contrasting tone categories, and the middle item (X) was the same tone as the first (A) or third (B) item; interstimulus interval (ISI) was 1 s. Listeners were asked to

click on the "1" or "3" displayed on the screen to indicate whether the X item matched category A or B. To avoid simple acoustic identity judgements, the X item was a different token of the same tone category as the matching A or B item. For each contrast, the four AXB trial types (AAB, ABB, BBA, BAA) occurred four times, and each of the four tokens per stimulus set occurred twice in each position (first, second, or third). There were 64 (4 syllables × 4 AXB trial types × 4 times) randomized trials for each of the six tone contrast blocks, resulting in a total of 384 (6 tone contrasts × 64 trials) trials.

Following a short break, participants then completed the tone categorization task using four tone icons displayed in the four quadrants of their screens. First came eight practice items (two tokens for each tone category) in random order, followed by the 64 (16 words × 4 tokens) trials of individual test items. The response time-out was 3.5 s. Participants were told to click a button in the centre of their screen to activate each trial, which was equidistant from the four tone contour icons, and one token of one of the 16 test pseudowords played out. Listeners then indicated which tone they had heard by clicking on one of the four tone icons. The tone icon positions were held constant for each participant and counterbalanced across participants. Listeners were instructed to hold the central activation button until they heard the stimulus. If they released it too soon, the trial ended automatically. Trials that ended automatically or lacked a response within 3.5 s (178 occurrences, or 2% of all trials) were repeated at the end of each block. Listeners were trained on the target words in consecutive six days with no more than two training sessions conducted in the same day. After completing the last training session, they completed the same tone discrimination and categorisation tests as during pre-training.

2.2. Results

2.2.1. Identification of Mandarin tones

Figure 1 shows mean percentage of choices for the four Mandarin tones in pre- and post-training tests by participants in the experimental (accent variability) and control (Beijing-only) word training conditions. Two criteria were used to determine tone assimilations [21] to the icons: (1) a given f_0 contour icon must be selected significantly more than chance level (25%), and (2) it must be chosen significantly more often than any other icons. For each group in each test phase (pre, post), separate one-sample *t*-tests against chance level 25% were

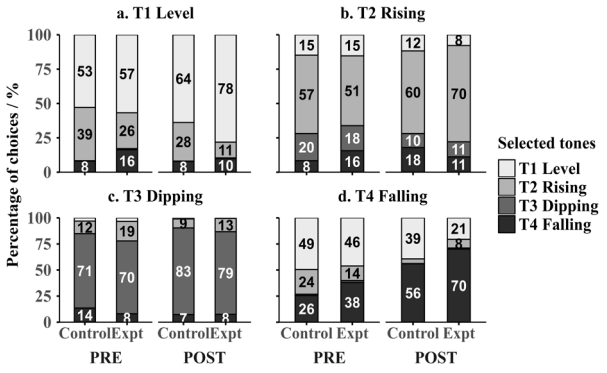

Figure 1: *Mean percentage of choices of the four tone icons for Mandarin tones in pre- and post-training tests by listeners in the experimental (accent variability) and control (Beijing-only) word training conditions.*

conducted for each tone to address criterion (1) using *R* with the *Student's t-Test* function. Multiple linear mixed-effects models were built for criterion (2) with the *lmer* function from package *lme4* [29]. Participants' percentage of choices for each icon was specified as the dependent variable. Training conditions, test phases, and the selected Mandarin tones were specified as fixed effects, and participants as a random effect. The Kenward-Roger approximation to the degrees of freedom was used to calculate the *p* values for the fixed-effects factors [30] and the *Anova* function from package *car* [31] was used to calculate *F*. Pairwise comparisons were conducted with *lsmeans* [32] in *R* whenever necessary to determine whether the percentage of choosing tone icon was greater than each of the other icons.

When responding to T1 level tone stimuli in the pre-training test, the listeners in both conditions split their choices between level (M_{expt} = 56.77%; $M_{control}$ = 52.86%) and rising (M_{expt} = 26.04%; $M_{control}$ = 38.80%) icons. Level icons were selected significantly above chance ($t(23) = 6.69$, $p < .001$) only in the experimental condition, and were selected significantly more often than the rising icons, Estimate = 30.73, $SE = 5.99$, $t(345) = 5.13$, $p < .0001$, indicating that these listeners Categorized T1 level tone before training. In the control condition, level and rising icons were both selected significantly above chance (T1: $t(23) = 5.44$, $p < .001$; T2: $t(23) = 2.77$, $p = .005$) and there was a non-significant difference between them (Estimate = 14.06, $SE = 5.99$, $t(345) = 2.34$, $p = .58$), indicating that T1 level tone stimuli were Uncategorized by the control participants. After training, for T1 stimuli only the level icon was selected significantly more than 25% by each group (M_{expt} = 78.12%, $t(23) = 9.49$, $p < .001$; $M_{control}$ = 63.08%, $t(23) = 5.37$, $p < .001$), suggesting that T1 level tone assimilation became Categorized for both groups. For responses to T2 rising and T3 dipping tone stimuli, listeners in both conditions selected rising and dipping icons, respectively, in both pre- and post-training tests above 50% which is significantly above chance, and each was chosen significantly more often than the other three icons, indicating that T2 rising and T3 dipping were correctly Categorized.

When responding to T4 falling tone stimuli, both groups split their choices between falling (M_{expt} = 38.02%; $M_{control}$ = 25.52%) and level (M_{expt} = 46.09%; $M_{control}$ = 49.47%) icons. Prior to training, participants in the experimental condition selected both falling ($t(23) = 2.51$, $p = .009$) and level ($t(23) = 4.25$, $p < .001$) icons significantly above chance and the difference of choices between them was non-significant (Estimate = 8.07, $SE = 6.42$, $t(345) = 1.25$, $p = 0.99$), indicating that T4 falling was Uncategorized. However, in the control condition only the level icon ($t(23) = 5.09$, $p < .001$) was selected significantly above chance, which was also selected significantly more often than the other three tone icons, indicating that the listeners Categorized T4 falling incorrectly to the level icon. After training, the experimental condition listeners categorized T4 falling tone as the falling icon. They selected only falling icons significantly more than chance (M = 70.05%; $t(23) = 5.98$, $p < .001$), which was also selected significantly more often than the other three tones, indicating that the listeners had correctly Categorized T4 as falling. While the control condition listeners improved by shifting part of their level-icon choices to falling icons, they still showed Uncategorized assimilation of T4 falling stimuli, because they chose both falling (M = 55.98%; $t(23) = 4.11$, $p < .001$) and level (M = 39.32%; $t(23) = 2.04$, $p = .02$) icons more often than chance, but with no significant difference between them (Estimate = -16.66, $SE = 6.42$, $t(345) = -2.59$, $p = .40$).

2.2.2. Discrimination of Mandarin tones

Figure 2 displays tone discrimination in both conditions for the pre- and post-training tests. To control for response bias, each participant's discrimination data were transformed to *A* scores [33]:

$$A = \begin{cases} \frac{3}{4} + \frac{H-F}{4} - F\,(1-H) & \text{if } F \le 0.5 \le H; \\ \frac{3}{4} + \frac{H-F}{4} - \frac{F}{4H} & \text{if } F \le H < 0.5; \\ \frac{3}{4} + \frac{H-F}{4} - \frac{1-H}{4(1-F)} & \text{if } 0.5 < F \le H \end{cases} \quad (1)$$

Where *F* is false alarm rate and *H* is hit rate (see [34] for details on Signal Detection Theory). Higher *A* values indicate better discrimination.

Multiple linear mixed-effects models were built on those values, with training conditions, test phases, and tone contrasts being specified as fixed effects and participants as a random effect. Calculation of *p* and *F* values was the same as in the tone categorization model. While there was no significant main effect of training conditions, the main effects of test phases, $F(1, 550) = 5.01$, $p = 0.02$, and tone contrasts, $F(5, 546) = 14.05$, $p < .001$, and the training conditions × test phases × tone contrasts interaction, $F(23, 528) = 3.67$, $p < .001$, were all significant. To tease the interaction apart, pairwise comparisons were conducted to assess improvement in tone discrimination between pre- and post-training tests for both groups.

Prior to training, while listeners in both groups were highly sensitive to tone differences, they showed significantly lower sensitivity to differences between T1 level vs. T4 falling ($M_{expt} = 0.91$ in *A* scores; $M_{control} = 0.90$) than between T1 level vs. T3 dipping ($M_{expt} = 0.97$, Estimate = 0.07, $SE = 0.02$, $t(528) = 4.24$, $p = .006$; $M_{control} = 0.98$, Estimate = 0.07, $SE = 0.02$, $t(528) = 4.61$, $p = .001$); and T2 rising vs. T3 dipping ($M_{expt} = 0.97$, Estimate = -0.06, $SE = 0.02$, $t(528) = -3.99$, $p = .02$; $M_{control} = 0.97$, Estimate = -0.07, $SE = 0.02$, $t(528) = -4.03$, $p = .01$); and T3 dipping vs. T4 falling ($M_{expt} = 0.97$, Estimate = -0.06, $SE = 0.02$, $t(528) = -3.85$, $p = .03$; $M_{control} = 0.96$, Estimate = -0.06, $SE = 0.02$, $t(528) = -3.80$, $p = .03$). *A* scores of tone pairs T1 level vs. T2 rising and T2 rising vs. T4 falling for both groups fell between those of tone pair T1 level vs. T4 falling and tone pairs involving T3 dipping. No significant differences between them were observed. While listeners in both groups showed no significant improvement of sensitivity to tone differences in each tone pair after training, differences among tone pairs were

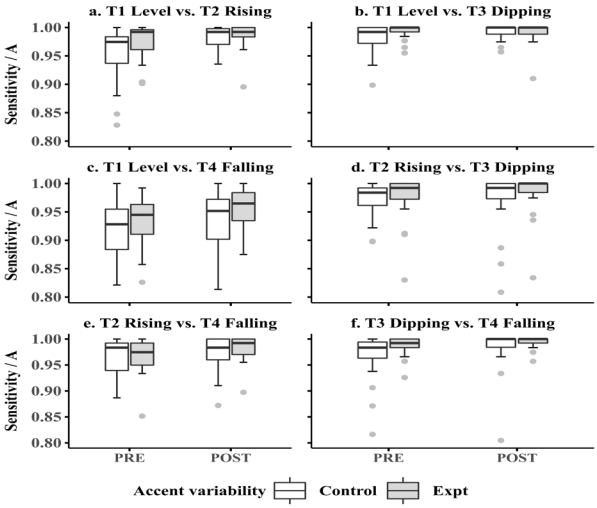

Figure 2: *AXB discrimination for the six tone contrasts in pre- and post-training tests by listeners in the experimental (accent variability) and control (Beijing-only) tone word training conditions. Outliers are displayed as grey points.*

no longer significant in the experimental condition, whereas listeners in the control condition continued to display less sensitivity to tone contrast T1 level vs. T4 falling ($M = 0.92$) than to T1 level vs. T3 dipping ($M = 0.97$; Estimate = 0.06, $SE = 0.02$, $t(528) = 3.55$, $p = .06$) and T3 dipping vs. T4 falling ($M = 0.98$; Estimate = -0.06, $SE = 0.02$, $t(528) = -3.86$, $p = .03$). Thus, only the experimental group improved on T1 level vs. T4 falling.

3. Discussion

This study investigated lexical tone contour categorization and discrimination by English listeners before and after minimal-tone-pair word training with only talker variability (control) or L2-Mandarin regional accent as well as talker variability (experimental). As predicted, accent variability facilitated post-training tone categorization and discrimination.

English listeners were required to assimilate Mandarin tones to tone contour icons due to the lack of lexical tones in their native phonological system. Before training, the experimental group showed Categorized assimilation of T1 level, which was Uncategorized by the control group. Both groups showed correctly Categorized T2 rising and T3 dipping, but Uncategorized or incorrectly Categorized assimilation of T4 falling. Their contrast assimilations were thus Two-Category for T1 level vs. T2 rising, T3 dipping, or T4 falling in experimental group, but T2 rising vs. T3 dipping for both groups. T1 level vs. T2 rising or T3 dipping in control group, T4 falling vs. T2 rising or T3 dipping in experimental group were Uncategorized-Categorized. T1 level vs. T4 falling was Single-Category for controls. Both groups discriminated all contrasts very well before training except for T1 level vs. T4 falling in control group, consistent with PAM predictions for these contrast assimilation types. Although lacking tones in their native phonological system, English listeners are *not* deaf to f_0 height and contour [5] as nonspeech patterns represented by icons. Moreover, minimal-tone-contrast word training with high accent and talker variability enhanced performance more than talker variability alone did.

After training, experimental condition Categorized all tones correctly, presumably as L2 phonological components. Correspondingly, their tone discriminations became equally excellent – their lower sensitivity to differences between T1 level and T4 falling than to other tone pairs prior to training disappeared after training. Listeners in the control group shifted their assimilation of T1 level from Uncategorized to Categorized. However, they failed to correctly Categorize T4 falling after training, although they did shift from incorrectly Categorizing it as T1 level to splitting their choices equally between falling and level icons, i.e., Uncategorized. Thus for them, T4 falling vs. T1 level became Uncategorized-Categorized after training. Their discrimination of this contrast remained poorer than for their Two-Category assimilation of T1 level vs. T3 dipping, again consistent with PAM principles.

Listeners in control condition improved their T1 level categorization to Categorized assimilation after minimal-tone-contrast word training, indicating that high talker variability alone can facilitate tone perception, which is in line with prior studies, such as[12], [15]. In addition to high talker variability, listeners in the experimental condition received high accent variability, which aided them forming T4 falling category. More importantly, their tone contour categorization and discrimination suggest that they had established L2 tone phonological categories for the tones, with excellent discrimination of all tone contrasts.

4. References

[1] Y. R. Chao, *Mandarin primer: An intensive course in spoken Chinese*. MA: Harvard University Press, 1948, p. 336.

[2] A. Abramson, "The tones of central Thai: Some perceptual experiments," in *Studies in Tai linguistics*, J. G. Harris and J. Chamberlain, Eds. Bangkok: Central Institute of English Language, 1975, pp. 1–16.

[3] C. K. So and C. T. Best, "Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties," *Poznan Studies in Contemporary Linguistics*, vol. 47, no. 1, pp. 133–145, 2011, doi: 10.2478/psicl-2011-0011.

[4] C. K. So and C. T. Best, "Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories," *Stud Second Lang Acquis*, vol. 36, no. 2, pp. 195–221, Jun. 2014, doi: 10.1017/S0272263114000047.

[5] P. A. Hallé, Y.-C. Chang, and C. T. Best, "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," *Journal of Phonetics*, vol. 32, pp. 395–421, 2004, doi: 10.1016/S0095-4470(03)00016-0.

[6] C. Kiriloff, "On the auditory perception of tones in Mandarin," *Phonetica*, vol. 20, pp. 63–67, 1969, doi: 10.1159/000259274.

[7] H. Bluhme and R. Burr, "An audio-visual display of pitch for teaching Chinese tones," *Stud. Linguistics*, vol. 22, pp. 51–57, 1971.

[8] X. S. Shen, "Toward a register approach in teaching Mandarin tones," *Journal of the Chinese Language Teachers Association*, vol. 24, pp. 27–47, 1989.

[9] C. K. So and C. T. Best, "Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences," *Language and Communication*, vol. 53, no. Pt 2, pp. 273–293, 2010, doi: 10.1111/j.1743-6109.2008.01122.x.Endothelial.

[10] Y. Wang, M. M. Spence, A. Jongman, and J. A. Sereno, "Training American listeners to perceive Mandarin tones," *The Journal of the Acoustical Society of America*, vol. 106, no. 6, pp. 3649–3658, 1999, doi: 10.1121/1.428217.

[11] K. Zhang, G. Peng, Y. Li, J. W. Minett, and W. S.-Y. Wang, "The effect of speech variability on tonal language speakers' second language lexical tone learning," *Frontiers in Psychology*, vol. 9, p. 1982, 2018, doi: 10.3389/fpsyg.2018.01982.

[12] S. Wiener, M. K. M. Chan, and K. Ito, "Do explicit instruction and high variability phonetic training improve nonnative speakers' Mandarin tone productions?," *The Modern Language Journal*, vol. 104, no. 1, pp. 152–168, 2020, doi: 10.1111/modl.12619.

[13] K. Johnson, "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," *Journal of Phonetics*, vol. 34, no. 4, pp. 485–499, 2006, doi: 10.1016/j.wocn.2005.08.004.

[14] J. Barcroft and M. S. Sommers, "Effects of acoustic variability on second language vocabulary learning," *Stud. Sec. Lang. Acq.*, vol. 27, no. 03, 2005, doi: 10.1017/S0272263105050175.

[15] P. C. M. Wong and T. K. Perrachione, "Learning pitch patterns in lexical identification by native English-speaking adults," *Applied Psycholinguistics*, vol. 28, no. 4, pp. 565–585, 2007, doi: 10.1017/S0142716407070312.

[16] T. Laméris and B. Post, "The combined effects of L1-specific and extralinguistic factors on individual performance in a tone categorization and word identification task by English-L1 and Mandarin-L1 speakers," *Second Language Research*, p. 026765832210900, Apr. 2022, doi: 10.1177/02676583221090068.

[17] Y. Li, C. T. Best, M. D. Tyler, and D. Burnham, "Regionally accented Mandarin lexical tones," *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2474–2475, 2020, doi: 10.1121/1.5146856.

[18] Y. Li, C. T. Best, M. D. Tyler, and D. Burnham, "Tone variations in regionally accented Mandarin," in *Proceedings of the 21st Annual Conference of the International Speech Communication Association*, 2020, pp. 4158–4162. doi: 10.21437/interspeech.2020-1235.

[19] 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.

[20] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 171–204.

[21] J. Chen, C. T. Best, and M. Antoniou, "Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners," *Journal of Phonetics*, vol. 83, p. 101013, 2020, doi: 10.1016/j.wocn.2020.101013.

[22] A. Reid *et al.*, "Perceptual assimilation of lexical tone: The roles of language experience and visual information," *Atten Percept Psychophys*, vol. 77, no. 2, pp. 571–591, 2015, doi: 10.3758/s13414-014-0791-3.

[23] A. R. Bowles, C. B. Chang, and V. P. Karuzis, "Pitch ability as an aptitude for tone learning: An aptitude for Tone," *Language Learning*, vol. 66, no. 4, pp. 774–808, Dec. 2016, doi: 10.1111/lang.12159.

[24] Q. Cai and M. Brysbaert, "SUBTLEX-CH: Chinese word and character frequencies based on film subtitles," *PLoS ONE*, vol. 5, no. 6, pp. e10729–e10729, 2010, doi: 10.1371/journal.pone.0010729.

[25] W. J. B. van Heuven, P. Mandera, E. Keuleers, and M. Brysbaert, "Subtlex-UK: A new and improved word frequency database for British English," *Quarterly Journal of Experimental Psychology*, vol. 67, no. 6, pp. 1176–1190, 2014, doi: 10.1080/17470218.2013.850521.

[26] R Core Team, *R: The R Project for Statistical Computing*. 2021. [Online]. Available: https://www.r-project.org/

[27] C. T. Best, G. W. McRoberts, and E. Goodell, "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," *The Journal of the Acoustical Society of America*, vol. 109, no. 2, pp. 775–794, 2001, doi: 10.1121/1.1332378.

[28] M. M. Faris, C. T. Best, and M. D. Tyler, "An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized," *The Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. EL1–EL5, 2016, doi: 10.1121/1.4939608.

[29] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *ArXiv e-prints*, vol. arXiv:1406, 2014, doi: 10.18637/jss.v067.i01.

[30] U. Halekoh and S. Højsgaard, "A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models-the R package pbkrtest," *Journal of Statistical Software*, vol. 59, no. 9, pp. 1–32, 2014.

[31] J. Fox and S. Weisberg, *An R Companion to Applied Regression*, 3rd edition. SAGE Publications, Inc, 2018.

[32] R. V. Lenth, "Least-squares means: The R package lsmeans," *Journal of Statistical Software*, vol. 69, no. 1, Art. no. 1, 2016, doi: 10.18637/jss.v069.i01.

[33] J. Zhang and S. T. Mueller, "A note on ROC analysis and non-parametric estimate of sensitivity," *Psychometrika*, vol. 70, no. 1, pp. 203–212, 2005, doi: 10.1007/s11336-003-1119-8.

[34] N. A. Macmillan and C. D. Creelman, *Detection theory: A user's guide*, 2nd ed. Mahwah, N.J.: Lawrence Erlbaum Associates, 2005.