INFLUENCE OF WITHIN-CATEGORY TONAL INFORMATION IN THE RECOGNITION OF MANDARIN-CHINESE WORDS BY NATIVE AND NON-NATIVE LISTENERS: AN EYE-TRACKING STUDY

Zhen Qin[1], Annie Tremblay[2], Jie Zhang[2]
[1]The Hong Kong Polytechnic University, Hong Kong; [2]University of Kansas, United States
zhen-quentin.qin@polyu.edu.hk, atrembla@ku.edu, zhang@ku.edu

ABSTRACT

Previous research showed that within-category phonetic details of segments constrain lexical activation. This study investigates how within-category tonal information influences native and non-native Mandarin listeners' spoken word recognition. Native Mandarin listeners and proficient English-speaking Mandarin learners were tested in a visual-world eye-tracking experiment. The target word contained a level tone and the competitor word contained a high-rising tone, or vice versa. The auditory stimuli were manipulated such that the target tone was either canonical (Standard), phonetically more distant from the competitor (Distant), or phonetically closer to the competitor (Close). Compared to the Standard condition, Mandarin listeners' target-over-competitor word activation was enhanced in the Distant condition and inhibited in the Close condition, whereas English listeners' target-over-competitor word activation was inhibited in both the Distant and Close conditions. These results suggest that within-category tonal information influences native and non-native Mandarin listeners' word recognition differently, explained by their different tonal processing mechanisms.

Keywords: Chinese tones, spoken word recognition, fine-grained tonal variability, eye tracking.

1. INTRODUCTION

The classic view of categorical perception assumes that the speech perception and processing mechanisms discard fine-grained phonetic variability to unmask the underlying phonemes [1]. This has been argued not only for the perception of consonants and vowels [2], but also for the perception of lexical tones in Mandarin Chinese (henceforth Mandarin) [3-4]. More recently, however, research on spoken word recognition that uses the visual-world eye-tracking paradigm [5] has shown that native listeners are in fact sensitive to the within-category phonetic variability in segments (e.g., VOT), with this information modulating target and competitor word activation (as indexed by listeners' fixations to target and competitor words) [6-8].What is unclear from previous research is whether within-category tonal variability also modulates word recognition in a language with contrastive tones such as Mandarin. The processing of Mandarin tones requires listeners to evaluate the pitch they hear against the pitch range of the talker, and continuously update this evaluation as more of the pitch contour is heard over time [9-10]. Importantly, a given Mandarin talker will show variability in the production of tonal categories, thus also requiring listeners to evaluate the pitch they hear against the talker's different realizations of the same tonal categories and of different tonal categories. Given the dynamic and variable nature of lexical tones, the question of whether within-category tonal variability influences word activation in Mandarin is not a trivial one.

The first goal of the current study is thus to test whether, and if so, how, fine-grained, within-category tonal variability modulates native Mandarin listeners' word recognition as the speech signal unfolds. If this fine-grained variability constrains lexical access, Mandarin listeners should show more competition from tonal competitors when the pitch of the target word (heard in the signal) is acoustically closer to that of the tonal competitor word (in listeners' lexical representation) than when it is acoustically more distant from that of the tonal competitor word.

Another unanswered question from the earlier research is whether second-language (L2) learners of Mandarin differ from native Mandarin listeners in their use of within-category tonal information. For instance, whether within-category tonal information constrains English listeners' recognition of Mandarin words is likely to depend on the degree to which English listeners' tone representations are phonetically detailed. L2 learners' difficulty in recognizing Mandarin tones has been attributed to their limited exposure to tonal variability, which may cause L2 learners' tonal representations to be coarser or less phonetically detailed, making it more difficult for them to interpret within-category tonal information in relation to prototypical tonal categories in lexical access [11-13].

The second goal of the current study is therefore to test whether English-speaking second language (L2) learners of Mandarin differ from Mandarin listeners in the use of within-category tonal variability in the recognition of Mandarin words. If L2 learners

of Mandarin are sensitive to the fine-grained phonetic details of lexical tones but have difficulty relating these details to prototypical tonal categories, they might show more competition from tonal competitors when the pitch of the target word (heard in the signal) is different from the standard (i.e., prototypical) tone, whether or not the pitch heard is acoustically closer to or more distant from that of the tonal competitor word.

2. METHODS

2.1. Participants

A total of 36 native Mandarin speakers and 26 proficient Mandarin learners who spoke English as their native language and learned Mandarin as L2 in college were recruited in Beijing, China. L2 learners' proficiency in Mandarin was tested with a Mandarin lexical decision task adapted from LexTALE [14], as well as a Mandarin cloze (i.e., fill-in-the-blank) test [15]. The L2 learners' language experience and mean percent proficiency scores are provided in Table 1.

Table 1: Means (and standard deviations, in parentheses) of learners' language background and proficiency variables.

Age of First Exposure (year)	17.6 (3.4)
Years of Mandarin Instruction	4.1 (2.3)
Length of Residence (month)	14 (13.3)
Lexical Decision Test (%)	66.5 (8)
Cloze Test (%)	82.1 (10)

2.2. Materials

Participants completed a visual-world eye-tracking experiment. All words in the experiment were imageable monosyllabic Mandarin nouns. In each trial, images corresponding to target, competitor, and distracter words appeared in the four cells of a (non-displayed) 2 x 2 grid (see Fig. 1). Six approximant-initial T1-T2 word pairs (e.g., T1, a level tone: /jā/ 'duck'; T2, a rising tone: /já/ 'tooth') were used as target and competitor words. The two words in each pair shared the same segments, contrasted in tones, and were not semantically related. When the target and competitor words carried T1 and T2, the two distracter words, which differed segmentally from the target and competitor words, carried T3 and T4 (e.g., T3, / tɕǐŋ/ 'well'; T4 /tɕìŋ/ 'mirror'). Which tone was contained in the target and competitor words was counter-balanced in two lists. The experiment included filler trials with T3-T4 target and competitor word pairs. The experiment included three repetitions for each auditory word and its corresponding display.

The auditory T1-T2 words had their pitch height resynthesized such that the pitch contour of the target word (heard in the signal) would be either more similar to or more different from that of the competitor word (in listeners' lexical representation). Three levels of a tonal continuum were created for each T1 and T2 (see Fig. 2). T1-standard and T2-standard are standard (S) exemplars of T1 and T2 natural tokens. T1-distant is acoustically more distant (D) from T2 than T1-S is, and T1-close is acoustically closer (C) to T2 than T1-S is. Likewise, T2-D is acoustically more distant from T1 than T2-S is, and T2-C is acoustically closer to T1 than T2-S is. The standard tokens were created by using the average pitch values of each the T1 and T2 tokens produced in isolation by a male native speaker of Mandarin. The acoustically closer and more distant tokens were created raising or lowering the starting average point of the contour by 10 Hz and interpolating between the new starting points and the natural endpoint. The resynthesized pitch contours (see Fig. 2) were superimposed on the duration-normalized stimuli.

Figure 1: A visual display of a T1-T2 trial used in the visual world paradigm (the orthographic transcriptions were not presented in the actual experiment).

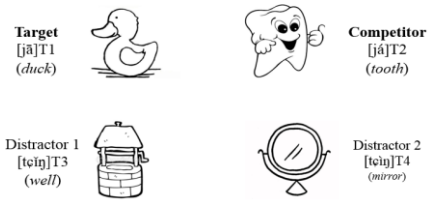
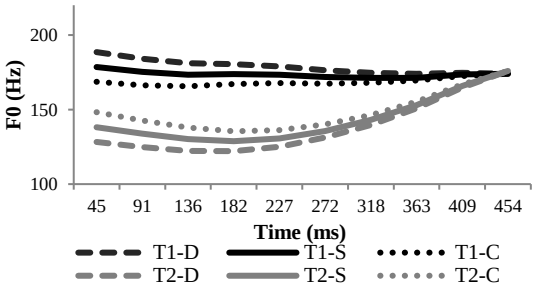


Figure 2: Tonal continua of Tone 1 (Level) and Tone 2 (Rising) used in the test trial.



2.3. Procedures

Participants first completed a word-picture association training so that they could distinguish the tonal contrasts and become familiar with the talker's pitch range. The training consisted of a picture selection test in which participants heard a spoken word and selected the picture corresponding to it from 20 candidate pictures (including the target, competitor, and distracter pictures). They received feedback on their responses, and the training ended when participants correctly identified all the pictures.

Participants then completed the eye-tracking experiment. They were instructed to click on the picture corresponding to the monosyllabic Mandarin word they heard through headphones. In each trial, participants first saw four pictures for two seconds

(preview phase). The pictures then disappeared and a fixation cross centered on the screen appeared and stayed on it for 500 ms. As the fixation cross disappeared, the four pictures reappeared and the auditory stimulus was simultaneously heard. Participants' eye movements were recorded from the onset of the auditory stimulus. Trials were split into four blocks, with each block containing only one repetition of each target word. The number of times each word and each tone was heard was counterbalanced within and across blocks. The experiment began with 12 practice trials.

To ease participants' memory burden, the training and experiment were conducted over three sessions, with participants learning a third of the word-picture associations and completing the corresponding portion of the experiment in each session.

2.4. Data analysis

Only correctly answered trials were analysed, resulting in the exclusion of 1% of the data for Mandarin listeners and 21.5% for English listeners. Proportions of fixations to the target, competitor, and distracter words were extracted in 20-ms time windows from the target word onset to 800 ms. The dependent variable for the statistical analyses was the difference between the empirical log-transformed proportions of target and competitor fixations from the onset to the offset of the target word with a 200-ms delay (i.e., from 200 to 654 ms) to accommodate the fact that eye movements take approximately 200 ms to reflect any processing of the speech signal [16].

Listeners' fixation differences were modeled using growth curve analysis (GCA) [17], a curvilinear regression that can model the linear, quadratic, and cubic shapes of the differential fixation lines. The GCAs were conducted with the *lme4* package in R [18]. For the sake of clarity, we present the analysis of each language group's results separately. These analyses included three orthogonal time polynomials (linear, quadratic, and cubic), condition (standard, distant, and close; baseline: standard), and the tone of the target word (T1 vs. T2; baseline: T1) as fixed effects. A backward-fitting function from the package *LMERConvenienceFunctions* in R [19] was used to identify the simplest model; only the results of the model with the best fit are presented, with *p* values being calculated with the *lmerTest* package in R [20]. All analyses included participant as random intercept and the orthogonal time polynomials as random slopes for the participant variable. Analyses yielding significant interactions between condition and tone were followed up by subsequent GCAs conducted separately for the T1 and T2, with the alpha level being adjusted to .025.

3. RESULTS

Figure 3 shows Mandarin and English listeners' differential proportions of fixations from the onset of the target word[i] to 800 ms. For both groups, the GCAs yielded significant tone-by-condition interactions.

Figure 3: Mandarin (top) and English (bottom) listeners' differential proportions of fixations in the standard, close, and distant conditions for T1 and T2; the shaded area represents one standard error above and below the mean; the vertical bars represent the beginning and end of the time window used for the statistical analyses

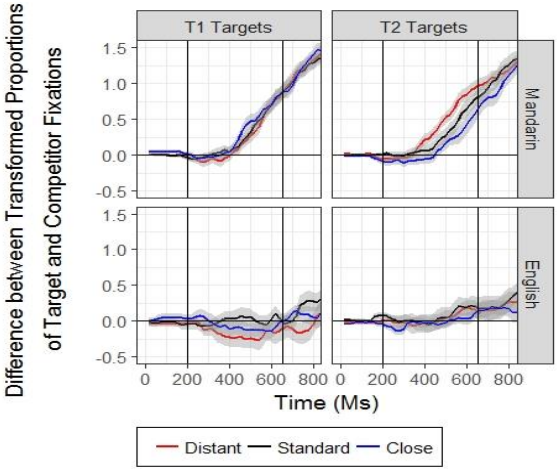


Table 2 presents the results of follow-up GCAs for Mandarin listeners. For T1, the best model did not include interactions between condition and time polynomials, suggesting that Mandarin listeners' word recognition was similar in all three conditions when the target contained a level tone. For T2, the significant positive estimate for the interaction between condition (distant) and the linear time polynomials suggests that Mandarin listeners had a fixation line with a steeper ascending slope in the distant condition than in the standard condition. The significant negative estimate for the interaction between condition (close) and the linear time polynomial indicates that Mandarin listeners' fixation line had a shallower ascending slope in the close condition than in the standard condition. These results suggest a decrease in tonal competition in the distant condition compared to the standard condition as a result of an acoustically greater tonal distance between the target and competitor, as well as an increase in tonal competition in the close condition compared to the standard condition as a result of an acoustically smaller tonal distance between the target and competitor. Mandarin listeners were thus sensitive to the within-category variability of the rising tone in their recognition of Mandarin words.

Table 2 also presents the results of follow-up GCAs for English listeners. For T1, the significant negative estimates for the interaction between

condition (distant; close) and the linear polynomial suggest that English listeners' fixation line had a shallower ascending slope in the distant condition or the close condition than in the standard condition. These results indicate that for T1, English listeners' word recognition was disadvantaged in both the distant and close conditions compared with the standard condition. For T2, the best model did not include any interactions between condition and the time polynomials, suggesting that English listeners' word recognition was similar in all three conditions.

Table 2: Growth curve analyses on the difference of Mandarin (top) and English (bottom) listeners' transformed target and competitor fixations for T1 and T2. α = .025; significant results in bold; D=distant; C=close.

Mandarin Listeners				
Tone	**Variable**	**Est.**	***t***	***p***
T1	Condition (D)	–0.043	–2.348	**.019**
	Condition (C)	0.034	1.829	.07
T2	Condition (D)	0.109	5.766	< **.001**
	Condition (C)	–0.132	–7.084	< **.001**
	Time × Condition (D) Linear	0.517	5.766	< **.001**
	Time × Condition (C) Linear	–0.200	–2.238	**.020**
English Listeners				
Tone	**Variable**	**Est.**	***t***	***p***
T1	Condition (D)	–0.159	–6.042	< **.001**
	Condition (C)	–0.059	–2.225	.027
	Time × Condition (D) Linear	–0.370	–2.930	< **.01**
	Time × Condition (C) Linear	–0.391	–3.097	< **.01**
T2	Condition (D)	–0.031	–1.198	.231
	Condition (C)	–0.083	–3.165	< **.01**

4. DISCUSSION AND CONCLUSION

This study investigated whether native Mandarin listeners' word recognition is constrained by within-category tonal variability, and whether native Mandarin listeners and English-speaking L2 learners of Mandarin differ in their use of this information. The results showed that native Mandarin listeners' target-over-competitor word activation was enhanced in the Distant condition and inhibited in the Close condition (compared to the Standard condition) for T2 target trials, whereas English listeners' target-over-competitor word activation was inhibited in both the Distant and Close conditions (compared to the Standard condition) for T1 target trials.

These findings indicate that Mandarin listeners do not discard the within-category phonetic details of T2, with small changes in the early pitch of the contour modulating listeners' word activation such that lexical representations are increasingly or decreasingly activated when the target tone is acoustically closer to or more distant from that of the competitor word. These findings provide further evidence that the speech processing system does not discard within-category tonal details, even when those details are not near the category boundary [6-8]. An important question that the current results raise, however, is why native Mandarin listeners showed sensitivity to within-category tonal information when it was heard in T2 targets but not when it was heard in T1 targets. Further research is needed to determine whether T1 stimuli that show a greater difference in Hz (e.g., 15 Hz) between the canonical and non-canonical tokens will be perceptually equivalent to T2 tokens and thus also modulate Mandarin listeners' word recognition.

Importantly, the results also showed that within-category tonal information *inhibited* English listeners' target-over-competitor word activation in trials with T1 targets. English listeners thus differed *qualitatively* from native Mandarin listeners in that within-category tonal information disrupted their word recognition independently of the phonetic distance between the target and competitor tones in trials with T1 targets. These difficulties may be due to L2 learners having coarser or less phonetically detailed representations of lexical tones as a result of their limited exposure to tonal variability [11-13]. Specifically, English listeners may not have been exposed to sufficient tonal input to develop robust representations of tones; as a result, they may not be able to organize the non-canonical tokens relative to the prototypical tokens in the tonal space, which may have increased tonal competition when the pitch contour of the target did not match the pitch contour expected for the prototypical tokens. Like with Mandarin listeners, the current results also raise the question of why English listeners showed an effect of within-category tonal information for only T1. Compared with T2, T1 has a stable pitch contour, and may stand out as the default tone for learners, as it is often the easiest tone to produce and perceive by non-native listeners [21-22]. These may explain why English listeners were more sensitive to the fine-grained variability of T1 compared to that of T2 [also see 23 for an asymmetry of T1-T2].

The present findings suggest that the native language has an important impact on listeners' processing of within-category tonal variability to resolve lexical competition in online word recognition. Future research with near-native L2 learners of Mandarin is necessary to investigate whether additional exposure to Mandarin would result in a more native-like performance.

5. ACKNOWLEDGEMENTS

This work was supported by a National Science Foundation Doctoral Dissertation Improvement Grant (BCS-1627554) and by the KU Doctoral Research Fund. Many thanks to: Dr. Xiaolin Zhou for providing us with lab access; Drs. Allard Jongman, Yan Li, James Magnuson, Stephen Politzer-Ahles, Joan Sereno, and the members of LING 850 at the University of Kansas for their helpful comments; and all of the participants for their time.

6. REFERENCES

[1] Stevens, K. N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am*. 111, 1872–1891.

[2] Liberman, A. M. et al. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*. 54, 358–368.

[3] Xu, Y. S., Gandour J., Francis J. 2006. Effects of language experience and stimulus complexity on the categorical perception of F0 direction. *J. Acoust. Soc. Am*. 120, 1063–1074.

[4] Peng, G. et al. 2010. The influence of language experience on categorical perception of F0 contour–levels. *Journal of Phonetics.* 38, 616–624.

[5] Tanenhaus, M. K. et al. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*. 268, 1632–1634.

[6] McMurray, B. et al. 2002. Gradient effects of within–category phonetic variation on lexical access. *Cognition.* 86, B33–B42.

[7] McMurray, B. et al. 2008. Tracking the timecourse of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin and Review.*15, 1064–1071.

[8] McMurray, B. 2009. Within–category VOT affects recovery from 'lexical' garden paths: Evidence against phoneme–level inhibition. *Journal of Memory and Language.* 60, 65–91.

[9] Malins, J. G., Joanisse, M. F. 2010. The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*. 62, 407–420.

[10] Shen, J. Deutsch, D. Rayner, K. 2013. On–line perception of Mandarin Tones 2 and 3: Evidence from eye movements. *J. Acoust. Soc. Am*. 133, 3016–3029.

[11] Díaz, B. et al. 2012. Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences.* 22, 680–689.

[12] Shen, G. Froud, K. 2018. Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: An ERP study. *Bilingualism: Language and Cognition.* 1–13.

[13] Qin, Z. Jongman, A. 2016. Does second language experience modulate perception of tones in a third language? *Language and Speech*. 59, 318–338.

[14] Lemhöfer, K. Broersma, M. 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods.* 44, 325–343.

[15] Yuan, B. 2009. Non–permanent representational deficit and apparent target–likeness in second language: Evidence from wh–words used as universal quantifiers in English and Japanese speakers' L2 Chinese. In: Snape, N., Leung, Y.–K. L., Sharwood Smith, M. (eds), *Representational Deficits in SLA: In honour of Roger Hawkins*. Amsterdam: John Benjamins Publishing. 69–103.

[16] Creel, S. C. 2014. Tipping the scales: auditory cue weighting changes over development. Journal of Experimental Psychology: Human Perception and Performance. 40, 1146–1160.

[17] Mirman, D. 2014. *Growth Curve Analysis and Visualization Using R.* Chapman and Hall / CRC.

[18] Bates, D. et al. 2015. Fitting linear mixed–effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

[19] Tremblay, A. Ransijn, J. 2015. *Model selection and post–hoc analysis for (G)LMER models*. Retrieved fromhttps://cran.rproject.org/web/packages/LMERConvenienceFunctions/.

[20] Kuznetsova, A.Brockhoff, B. Christensen, H. 2016. *Tests in linear mixed effects models*. Retrieved from https://cran.rproject.org/web/packages/lmerTest/index.html.

[21] Hao, Y. C. 2012. Second language acquisition of Mandarin Chinese tones by tonal and non–tonal language speakers. *Journal of Phonetics*. 40, 269–279.

[22] Wang, Y. et al. 1999. Training American listeners to perceive Mandarin tones. *J. Acoust. Soc. Am*. 106, 3649–3658.

[23] Shih, C., Lu, H. Y. 2015. Effects of Talker-to-Listener Distance on Tone. *Journal of Phonetics*. 51, 6–35. https://doi.org/10.1016/J.WOCN.2015.02.002.

[i] The time window of analysis used for the GCA was the onset and offset of the tonal portion of the stimuli, with a delay of 200 ms.