Effects of phonological and talker familiarity on second language lexical development

Jiang Liu and Seth Wiener
University of South Carolina | Carnegie Mellon University

Previous research has shown that second language (L2) learners of Mandarin learn new words more easily if the new word is homophonous with a word they already know (Liu and Wiener, 2020). That research involved word learning in which speech was produced by a single-talker with a specific pitch range. The present study examines whether the observed tonal homophone advantage is dependent on familiarity with the talker. Adult learners of Mandarin Chinese as an L2 were taught 20 new tonal words for three consecutive days. To manipulate phonological familiarity, 10 words had homophones already known to the learners and 10 words did not. To manipulate talker familiarity, participants were trained on a single talker but tested on 16 new talkers or trained and tested on 16 (multi)-talkers. Daily testing involved a 4-alternative-force-choice task. Both groups showed increased accuracy and faster response times on Day 2 compared to Day 1, but this learning was independent of homophone status or talker group. No other effects were found. These results suggest that the tonal homophone advantage in L2 word learning observed by Liu and Wiener (2020) may have been partially driven by an exceptionally high level of talker familiarity, since that study used a single speaker both for training and testing.

Keywords: lexical development, phonological learning, acoustic variability, second language acquisition

Lexical development in a second language has attracted considerable attention in the last twenty years (e.g., Dong, Gui, & MacWhinney, 2005; Wolter, 2001; Zareva, 2007). Among this body of work is Jiang's (2000, 2018) psycholinguistic model of lexical representations in which first language (L1) and second language (L2) words consist of a lemma (i.e., the grammatical form of a representation) and lexeme (i.e., the pairing of the form to its meaning, which can involve phonology

https://doi.org/10.1075/ml.20024.liu | Published online: 18 March 2022
The Mental Lexicon 17:1 (2022), pp. 132–153. ISSN 1871-1340 | E-ISSN 1871-1375

and/or orthography; see Levelt, 1989, 1993). Important to Jiang's model is the claim that the learning process influences how L1 and L2 lexical representations are formed. Children acquire all the components in an L1 lexical entry subconsciously through highly contextualized input and without connections to other languages. Adults, however, acquire the components involved in an L2 lexical entry consciously and typically without conceptual development.

Jiang proposes three stages of L2 lexical development. The first or formal stage of an L2 lexical entry involves only phonological and orthographical information. During this stage, the L2 representation relies on the L1 lexical entry for conceptual information through translation. Next, L1 lemma information is copied to the L2 lexical entry. During this lemma mediation stage, L2 learners are often observed using L2 vocabulary with incorrect L1 grammatical specifications (e.g., Wolter, 2006). Experimental evidence from L1-L2 translation studies like Talamas et al. (1999) have documented this gradual shift in reliance on L1 form and meaning. Talamas et al. presented L2-L1 word pairs and asked participants to indicate whether the pairs were correct translations of one another (e.g., Spanish *casa* – house). Incorrect translations that shared form with the correct translation were shown (e.g., *casa* – hound) as were incorrect translations that shared semantics with the correct translation (e.g., *casa* – apartment). Less proficient L2 learners showed more interference from incorrect form trials whereas more proficient L2 learners showed more interference from incorrect semantic trials (see also Sundermen & Kroll, 2006; Wiener & Tokowicz, 2021). During the third and final stage the L2 lemma is fully formed and L2 learners are able to use the word more accurately and appropriately (Kroll, Michael, Tokowicz, & Dufour, 2002; Talamas, Kroll, & Dufour, 1999). Jiang notes, however, that many L2 words may stop short of this final stage as learners continue to rely on L1 lemma mediation.

Jiang (2000, 2002, 2018) therefore characterizes L2 lexical development by its conscious and deliberate learning of form-meaning connections. Here we examine the role of phonological and talker familiarity on this process. Theoretically, how L2 lexical representations are formed and subsequently accessed depend on the match between the representation (corresponding to the phonological sound pattern) and the signal (produced by a specific talker). Familiarity with one – or both – sources of information may have an immediate effect on L2 lexical processes. We explore this hypothesis using Mandarin Chinese as our test language. Standard Mandarin Chinese (hereafter 'Mandarin') has a relatively small number of unmarked (consonant)-vowel-(nasal) syllables like *shu* (Duanmu, 2007). Many (though not all) of the roughly 400 unique Mandarin syllable types can stand alone as a morpheme or word (Chao, 1968; Packard, 2000; Zhou & Marslen-Wilson, 1995). Spoken word meaning is indicated by consonant and vowel segments along with one of four pitch patterns or tones on

the vowel. The syllable *shu* spoken with a high-level pitch (Tone 1) can mean 'uncle,' but the same syllable *shu* spoken with a low-dipping pitch (Tone 3) can mean 'mouse.' Because Chinese writing is morphemic in nature, a monosyllabic syllable+tone combination like *shu1* can mean 'uncle,' (叔) 'book,' (书) 'to lose,' (输) 'vegetable,' (蔬) and nearly 20 other homophonous *shu1* morphemes/words. On average, each syllable+tone combination represents 11 semantically and orthographically unique homophone mates (Tan & Perfetti, 1998). For L2 listeners, Mandarin word learning involves perceiving lexical tone in the face of high phonological overlap and homophony. In this study, we capitalize on these linguistic features to examine the role of phonological and talker familiarity on lexical development in adult learners of Mandarin as a second language.

Phonological overlap and L2 lexical development

Phonological overlap between known words and novel words facilitates word learning, recall, and production for L1 speakers. Words with many similar sounding neighbors (e.g., 'pear' is similar to 'care,' 'hair,' 'bear,' etc.) are learned faster, more accurately, and more completely than words that have few or no similar sounding neighbors, e.g., 'orange' (Storkel, 2001; Vitevitch, 2002; Vitevitch & Storkel, 2013). Similar facilitative neighborhood density effects (Luce & Pisoni, 1998) have been observed for L2 learners (e.g., Bradlow & Pisoni, 1999; Storkel, Armbrüster & Hogan, 2006; Wilcox & Medina, 2013). Stamer and Vitevitch (2012) taught L1 English-L2 Spanish adult learners 16 new Spanish words. Half of the words had dense phonological neighborhoods while the other half had sparse phonological neighborhoods. The authors tested participants both immediately after training and 48–72 hours later, which allowed lexical integration and overnight consolidation to occur (e.g., Dumay & Gaskell, 2007; Elgort, 2011). Testing involved a picture-naming task, a referent identification task, and a perceptual identification in noise task. In all three tasks, the L2 learners demonstrated a learning advantage for words that sounded similar to many Spanish words already known by the learners.

More recently, Liu and Wiener (2020) demonstrated that L2 learners can also use suprasegmental overlap to bootstrap lexical development. The authors taught L1 English-L2 Mandarin adult learners 20 new words for three consecutive days. Ten of the words had homophones that were already known to the learners. For example, participants had previously learned that *shu1* means 'book' but not that it could also mean 'uncle,' which served as the target word. The other 10 words had no homophones that the participant had already learned. For instance, the target word *shu3* 'mouse' was the only *shu3* word that the participants had encoun-

tered at that point in their L2 learning. Each day after training on the 20 words, participants completed a 4-alternative-forced-choice (4-AFC) task, which showed two images that shared the same syllable but differed in tone (e.g., *shu1* and *shu3*) along with two distractor images that differed in syllable and tone. These two distractor images came from the 20 words taught to the participants and always consisted of one word that had homophones that were already known to the learners and one word that did not. Participants were asked to click on the image that matched the perceived spoken word as quickly and accurately as possible.

The effect of homophony was not consistent across the three tests: on Day 1, there was an effect of homophony on accuracy but not speed, whereas on Day 2 and Day 3, there was an effect on speed but not accuracy. On Day 1, words with homophones already known to the learners (e.g., *shu1*) were identified more accurately than words that had no already known homophones (e.g., *shu3*). The authors claimed that this was because learners had previously acquired a phonological representation, which led to rapid integration into the lexicon. On Day 2 and Day 3, no accuracy difference was observed, suggesting overnight consolidation of tones (e.g., Dumay & Gaskell, 2007; Qin & Zhang, 2019) diminished this effect. On Day 2 and Day 3 participants identified new words with homophones already known to the learners (e.g., *shu1*) faster than new words that had no already known homophones (e.g., *shu3*). This response time difference suggested that once the new words were integrated or configured into the L2 lexicon and engaged with other words (e.g., Leach & Samuel, 2007), learners were able to access these new representations faster given the familiar phonology.

The present study extends Liu and Wiener's (2020) findings in two ways. First, we clarify whether L2 learners of Mandarin only benefit from complete phonological overlap (as in the case of tonal homophones such as *shu1* 'book' and *shu1* 'uncle') or partial phonological overlap (as in the case of knowing many words with 'u' vowels and a high-level tone like *qu1, du1, fu1*). The observed homophone effect in Liu and Wiener may have actually been a phonological neighbor effect driven by certain items having more neighbors than other items. Second, we clarify whether the homophone effect was driven by familiarity with the talker. Can adult L2 learners still take advantage of phonological overlap between known and new words when listening to unfamiliar talkers with variable segmental and tonal productions?

Talker variability and L2 lexical processing

The spoken input L1 and L2 learners receive is noticeably different. L1 acquisition typically involves speech from multiple caretakers, including highly variable

"motherese," which contains acoustically exaggerated cues (e.g., Kuhl, 2004; Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989). In contrast, L2 acquisition, particularly within a structured classroom setting, typically involves speech from one language instructor, which limits learners' exposure to acoustically variable speech cues (see Colantoni, Steele, Escudero, & Neyra, 2015; Flege, 1995).

Previous research has examined the variability of a particular talker in terms of indexical information or talker-dependent cues (Sommers & Barcroft, 2007, 2011) and distributional information or talker-independent cues (Escudero, Benders, & Wanrooij, 2011). L2 learners are sensitive to talker-dependent and talker-independent cues, benefit from acoustically variable multi-talker input in terms of phonetic categorization and discrimination, and are able to generalize to new talkers and new phonological contexts (e.g., Barcroft & Sommers, 2005, 2014; Hardison, 2003, 2005; Lively, Logan & Pisoni, 1993; though see Wade, Jongman, & Sereno, 2007 for conflicting results). However, multi-talker speech has been found to negatively affect L2 spoken word recognition. For most L2 learners, acoustically varied input from unfamiliar talkers increases lexical competition and often causes delayed or inaccurate word recognition (Kroll et al., 2002; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Weber & Cutler, 2004).

With respect to Mandarin, talker variability poses an interesting problem (see Pelzl, 2019 for a review). Lexical tone is primarily indicated by a talker's fundamental frequency (F0; Ho, 1976), yet F0 information varies as a function of a talker's age and sex. A listener must therefore filter out or normalize a talker's voice to identify relevant phonological tone category information (Leather, 1983; Moore & Jongman, 1997). Figure 1 plots the loess-smoothed F0 contours of 320 utterances spoken by 16 talkers (8 female; 8 male plotted in gray) along with overall mean female and male contours (plotted in color online). This figure highlights how tone productions vary considerably across talkers. For example, a female's phonologically low-dipping tone (Tone 3) and a male's phonologically high-level tone (Tone 1) may demonstrate similar F0 contours. This could lead to uncertainty over whether an unfamiliar talker saw her uncle (*shu1*) or a mouse (*shu3*).

How does talker variability affect L1 English-L2 Mandarin learners? In terms of L2 tone category learning, multi-talker input appears to be beneficial to adult learners as it promotes robust categories. Wang, Spence, Jongman, and Sereno (1999) trained L1 English-L2 Mandarin classroom learners for two weeks on tones produced by four talkers. A pre/post forced-choice tone categorization task was used to assess training improvements. On average, participants improved by roughly 20% and showed similar improvements to new stimuli and new talkers. Follow-up studies, however, suggested that individual learner differences may have affected the results (Dong, Clayards, Brown, & Wonnacott, 2019; Lee, Tao, & Bond, 2009, 2013; Perrachione, Lee, Ha, & Wong, 2011; Sadakata & McQueen, 2014).

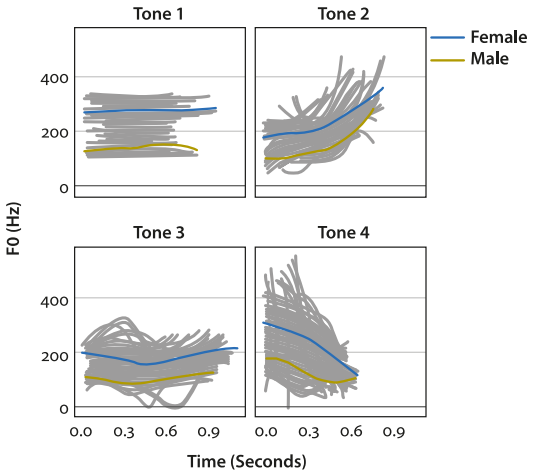

Figure 1. Loess-smoothed tone contours produced by 8 female and 8 male talkers along with female and male mean contours

In terms of L2 Mandarin spoken word recognition, Wiener, Ito, and Speer (2018, 2021) recorded eye movements as classroom L2 Mandarin learners and naïve monolingual English speakers were trained and tested on an artificial Mandarin-like language. After four consecutive days of sound-image learning, participants performed a 4-AFC task involving a target and tonal competitor (and two distractors). Participants trained and tested on multi-talker speech showed increased competition between the target and a tonal competitor relative to those trained and tested on the same talker. Evidence from the gating paradigm extended this work using natural Mandarin speech. Adult classroom L2 Mandarin learners heard speech from either a single talker (Wiener, Lee, & Tao, 2019) or 16 different talkers (Wiener & Lee, 2020) and were asked to report the perceived word given increasing longer fragments or gates of speech. On average, the L2 listeners demonstrated reduced spoken word recognition accuracy when listening to multi-talker input compared to single-talker input, though large variation was observed across the learners in line with previous individual differences studies at the tone and word level (e.g., Chandrasekaran et al. 2010; Perrachione et al. 2011; Wong & Perrachione, 2007).

In sum, previous L2 Mandarin findings indicate that increased phonological overlap in the form of syllable+tone homophones is advantageous for L2 word learning, so long as the speaker is familiar to the listener (Liu and Wiener, 2020). At the same time, increased talker variability in the form of input from multiple unfamiliar talkers is disadvantageous for L2 word recognition as it increases lexical competition and often results in inaccurate word recognition (Lee and Wiener, 2020; Wiener et al., 2018; Wiener and Lee, 2020).

The present study

The present study tests whether adult L2 learners can still take advantage of phonological overlap between known and new words when being trained and tested on unfamiliar talkers with variable F0 ranges. We train and test participants in two groups defined by their exposure to different talkers. The 'single-talker' group is trained on the target words from the same female talker used in Liu and Wiener (2020). This group is then tested on 16 unfamiliar talkers. This allows us to compare the accuracy and recognition response time given a familiar single-talker (analogous to a language instructor in a classroom) but wholly new unfamiliar interlocuters. In contrast, the 'multi-talker' group is both trained and tested on the target words from 16 different talkers. This allows us to compare the accuracy and recognition response time given multiple talkers at both training and testing – a scenario potentially more like L1 acquisition in which constant talker normalization is required by the listener (Kuhl, 2004). If phonological overlap is advantageous to learners irrespective of talker familiarity and the variability in the speech signal, then both groups should show increased accuracy for words with homophones already known to the learners compared to words without known homophones. That is, the accuracy results should align with Liu and Wiener's (2020) results despite the different, unfamiliar talkers and high variability stimuli used in the present study. If phonological overlap is only advantageous when the talker is familiar to the listener, then the multi-talker group should recognize homophones more accurately than the single-talker group, given that the latter group is trained on only one talker but tested on 16 different talkers whereas the multi-talker group is at least exposed to all 16 talkers in the training. In both instances, the correct response times may be delayed for the two groups given that increased variability from multiple talkers causes more tonal competition in L2 listeners (Wiener et al., 2018; Wiener & Lee, 2020). Finally, if any homophone advantage is found, we clarify whether the advantage is due to complete phonological overlap involving segments and tones, or partial overlap involving the onset-only, vowel-only, tone-only, onset+tone, or vowel+tone.

Method

Participants

39 native English speakers (20 male; 19 female; mean age = 18; SD = 0.7; age range: 18–22) participated in the study. All participants were enrolled in an in-person second semester Mandarin class at a public U.S. university. Although participants

were recruited from multiple classes with different meeting times, all participants used the same classroom materials and had the same instructors. No participant self-reported speaking an additional language. All participants started to learn Mandarin from college and had completed roughly 20 weeks of formal classroom instruction at the time of the experiment. All participants self-reported normal hearing and normal or corrected-to-normal vision. All participants received extra credit and a small payment for their participation.

Materials

Stimuli were taken from Liu and Wiener (2020), which consisted of 10 distinct consonant-vowel-(consonant) syllables, each paired with two different tones (all material and data available on OpenScience: https://osf.io/hycw2/). For each syllable, one tone served as a homophone based on a pre-test vocabulary screening while another tone served as a non-homophone.

Eight female and eight male native Mandarin speakers recorded the 20 target words in a sound-attenuated booth. Recordings were saved at 44k Hz/16 bits using Praat (Boersma & Weenink, 2018). Stimuli were clearly enunciated to promote L2 speech learning following Escudero et al. (2011). Each word was paired with a color drawing designed to establish the sound and meaning association (see Figure 2; online for color version).

Procedure

Training and testing occurred in a quiet lab for three consecutive days. The 39 participants were randomly split across the two groups (19 were trained on a single-talker, 9F, 10M; 20 were trained on 16 different talkers, 10F, 10M). Because pitch perception ability affects tone learning (Bowles, Chang, & Karuzis, 2016), on Day 1 all participants performed a pitch screening task (Tonometric: Mandell, 2018), which measured the smallest difference in Hz at which the listener could discriminate between two pure tones. Whereas any cutoff is relatively arbitrary, we removed participants who were unable to reliably discriminate between two pure tones at 30 Hz (or greater). This served to remove participants with potential congenital amusia (see Zhu et al., 2022) while still allowing us to examine a large enough dataset for reliable statistics. Two participants in the single-talker group were removed due to poor pitch perception abilities. This left 37 participants. The two groups did not differ in their pitch discrimination abilities (single-talker group mean of 9.7 Hz; multi-talker group mean of 7.1 Hz; $F(1,35) = 1.03$, $p = .32$).

On each day, participants performed a self-paced passive listening task (hereafter 'training') in which a spoken word and its corresponding image were simul-

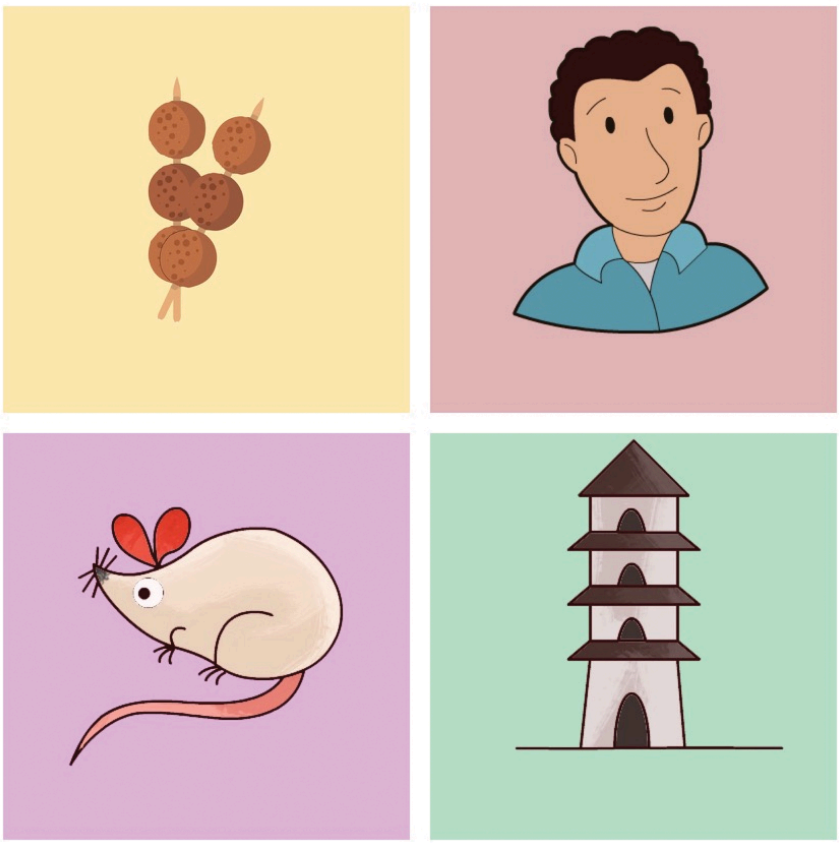

Figure 2. Example 4-AFC slide. Bottom left: *shu3* 'mouse' (target); top left: *wan2* '(meat) ball' (distractor); top right: *shu1* 'uncle' (tonal competitor); bottom right: *ta3* 'pagoda' (distractor)

taneously presented via computer and headphones. Participants were told to remember the sound-image pair and to mouse-click to advance to the next trial. Each word was presented four times in a pseudo-randomized order (80 total trials). This number was determined through piloting, which attempted to capture a Day 1 accuracy above chance (.25) but well below ceiling. After the training, participants were tested on the 20 learned words using a 4-AFC word identification task. During each trial, four images from the 20 trained words were presented on a computer monitor while one of the image's audio labels, i.e., the target, was presented over headphones. Participants were told to mouse-click on the image corresponding to the perceived audio as accurately and quickly as possible. In addition to the target, a tonal competitor and two distractors taken from the 20 trained words were displayed. Across all three days of testing, the positions of the target, tonal competitor, and distractors were counterbalanced, which resulted in 60 unique 2×2 image displays or 20 different slides per day. Figure 2 shows an example with *shu3* 'mouse' as the target and *shu1* 'uncle' as

the tonal competitor. This meant each of the 20 words was heard five times (four times in training and one time in testing) on each day. Corrective feedback on the first two days of 4-AFC testing guided learning. After mouse-clicking in the 4-AFC task, a blue circle appeared on the target. An additional mouse-click advanced the experiment to the next trial with a 1-second interval. All tasks were presented using E-prime 2.0 (Psychological Software Tools Inc.). After completing the 4-AFC task on the third day, participants performed a 5-minute naming task (not reported here). In total, the training and testing tasks took approximately 45 minutes each day.

Whereas the tasks and procedure were the same for all the participants, the audio heard during the training differed by group. The single-talker participants were trained on one female talker, but tested each day in the 4-AFC task on 16 different talkers. Across the three days of testing, no utterance was repeated for a total of 60 unique exemplars. The multi-talker participants were trained and tested on 16 different talkers. Across the three days of training and testing, no utterance was repeated for a total of 300 different exemplars [(80 training + 20 4-AFC testing) * 3 days)] of the possible 320 (20 items * 16 talkers). In other words, both groups always heard 20 new utterances from 16 different talkers during the 4-AFC testing. Figure 1 plots all 320 utterances to highlight the talker variability involved in tone production.

Data analysis

We adopt a .05 alpha-level for our analyses. 4-AFC mouse click accuracy results were analyzed using mixed-effects logistic regression modeling with the *lme4* package in R (version 3.6.2; R Core Team, 2019) and the "bobyqa" optimizer. The model contained mouse click accuracy (1 correct; 0 incorrect) as the outcome variable. Fixed effects included Talker group, Homophone status, and Day all as treatment coded variables. The single-talker group served as the reference level for talker group. A positive coefficient for the multi-talker group will indicate an increase in accuracy relative to the single-talker group. Words with already learned homophones served as the reference level for homophone status. A positive coefficient for the non-homophone status will indicate an increase in accuracy for words without already learned homophones. Day 2 served as the reference level for day (allowing for two comparisons: Day 1 vs. Day 2; Day 2 vs. Day 3). By making Day 2 the reference level, the expected coefficients for Day 1 and Day 3 should go in opposite directions. In other words, a positive coefficient for Day 3 will indicate an improvement in accuracy from Day 2 to Day 3, whereas a positive coefficient for Day 1 will indicate a decline in accuracy from Day 1 to Day 2. These three fixed effects and all corresponding two-way and three-way interac-

tions were included in the full confirmatory model. Random item and participant intercepts were included.

Correct response times (RTs) were analyzed using a mixed-effects linear regression model with the same variables, coding, and model structure as the accuracy model. We first treated correct RTs slower than 5000 ms as outliers and removed them. Given the nature of the sound-image mapping task, we believed responses beyond 5000 ms were largely uninformative (see Jiang, 2013 for discussion). From this subset, RTs beyond three standard deviations from the mean were removed. This filtering process removed 30% of the RT data, leaving 1,176 observations. In both models, high variance inflation factors (>10) were found for all interactions and therefore removed from the final model.

Results

4-AFC mouse click accuracy

Figure 3 (color online) plots the individual daily performance of each participant (points with jitter added), along with Talker group density, group means (solid lines), and group 95% confidence intervals (white box) for both the Homophone and Non-homophone conditions. The plot shows a general trend of overall improvement, with the two homophone conditions showing roughly similar accuracy. Tone mistakes made up of 90% of the errors with only 10% of the mistakes coming from an incorrect mouse-click on a distractor. There was no relationship between type of mistake and homophone status ($\chi^2(1) = 2.33$, $p = .13$), type of mistake and talker group ($\chi^2(1) = 0.52$, $p = .47$), or type of mistake and day ($\chi^2(2) = 4.40$, $p = .11$).

Table 1 reports the summary of the exploratory logistic regression model. Figure 4 visualizes the fixed effects' coefficients from the exploratory model. The model revealed a simple effect of Day. Day 2 accuracy (mean = .79) was higher than Day 1 accuracy (mean = .67). Day 3 accuracy (mean = .83) did not differ from Day 2 accuracy. All other fixed effects were null.

Correct 4-AFC mouse-click response times

Figure 5 (color online) plots the individual daily performance of each participant along with group results following the same plotting scheme as Figure 3. Table 2 reports the summary of the exploratory logistic regression model. Figure 6 visualizes the fixed effects' coefficients from the exploratory model. The model revealed a simple effect of Day. Day 2 RT (mean = 1826 ms) was faster than Day 1 RT

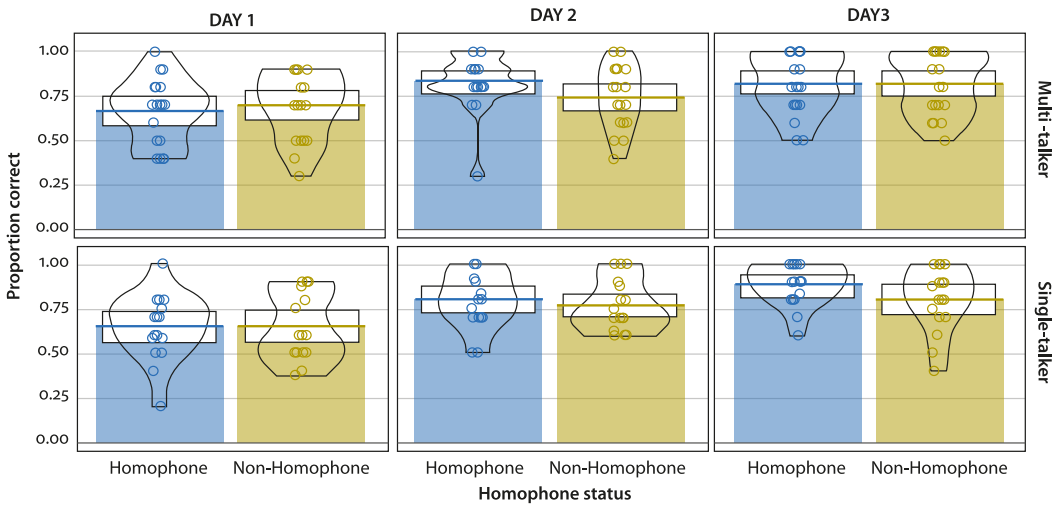

Figure 3. 4-AFC accuracy by participant (point), Talker variability (color), Homophone status, and Day. White boxes represent 95% confidence intervals with solid color line representing mean

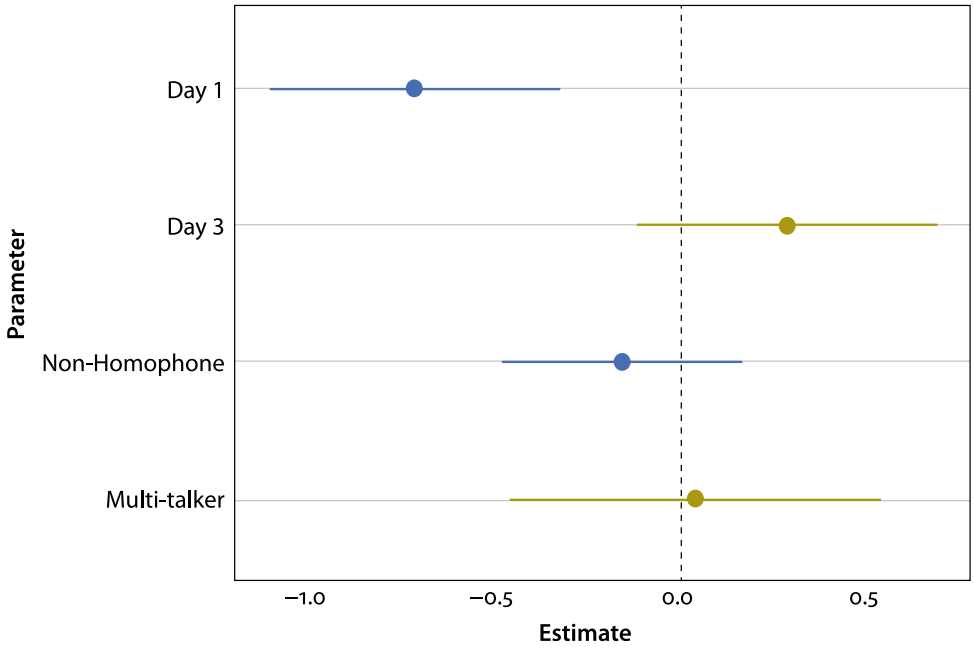

Figure 4. Estimates from mixed-effect logistic regression model predicting 4-AFC accuracy. Blue indicates a negative effect whereas yellow indicates a positive effect

(mean = 1906 ms). Day 3 RT (mean = 1860 ms) did not differ from Day 2 RT. All other fixed effects were null.

Table 1. Summary of mixed-effects logistic regression model predicting 4-AFC accuracy

Parameter	Log-Odds	SE	95% CI	*z*	*p*
(Intercept)	1.58	0.25	[1.09, 2.07]	6.34	< .001
Day 1	−0.73	0.20	[−1.12, −0.33]	−3.60	< .001
Day 3	0.29	0.21	[−0.12, 0.70]	1.38	0.166
Non-Homophone	−0.16	0.17	[−0.49, 0.17]	−0.95	0.341
Multi-talker	0.04	0.26	[−0.47, 0.54]	0.15	0.885
Random effects	**Variance**	**Sth. Deviation**			
Participant	0.50	0.71			
Item	0.24	0.49			

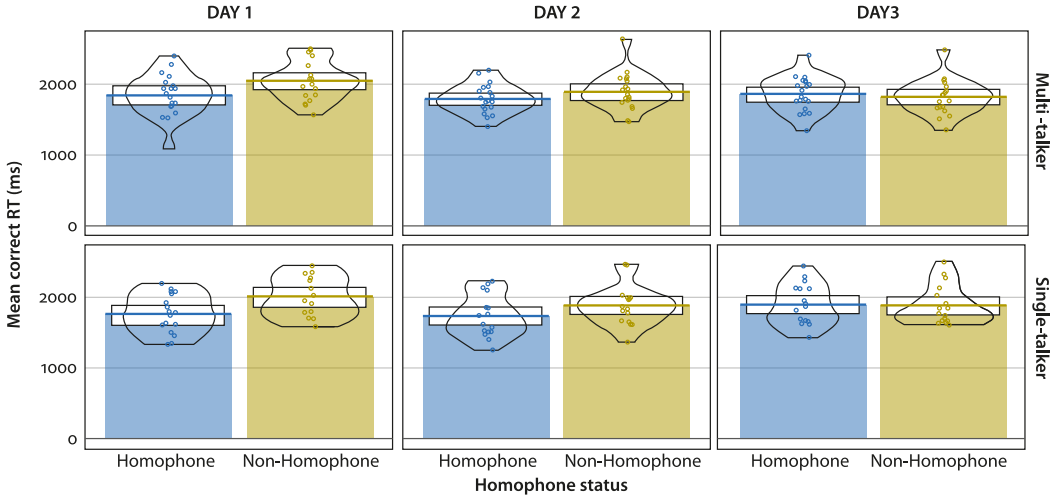

Figure 5. Correct 4-AFC mouse-click response times by participant (point), Talker variability (color), Homophone status, and Day. White boxes represent 95% confidence intervals with solid color line representing mean

Discussion

In this study, we set out to explore how familiarity with the signal-independent phonology of the sound-based structure of the lexicon (e.g., Leach & Samuel, 2007; Lindblom, 1990; Storkel et al., 2006; Stamer & Vitevitch, 2012) and familiarity with the signal-dependent talker information of voice and articulatory patterns (e.g., Bradlow & Pisoni, 1999; Lee et al., 2009; Nygaard et al., 1994) jointly affected L2 lexical learning and access of syllable+tone words. This study served as a follow-up to Liu and Wiener (2020). We asked whether adult L2

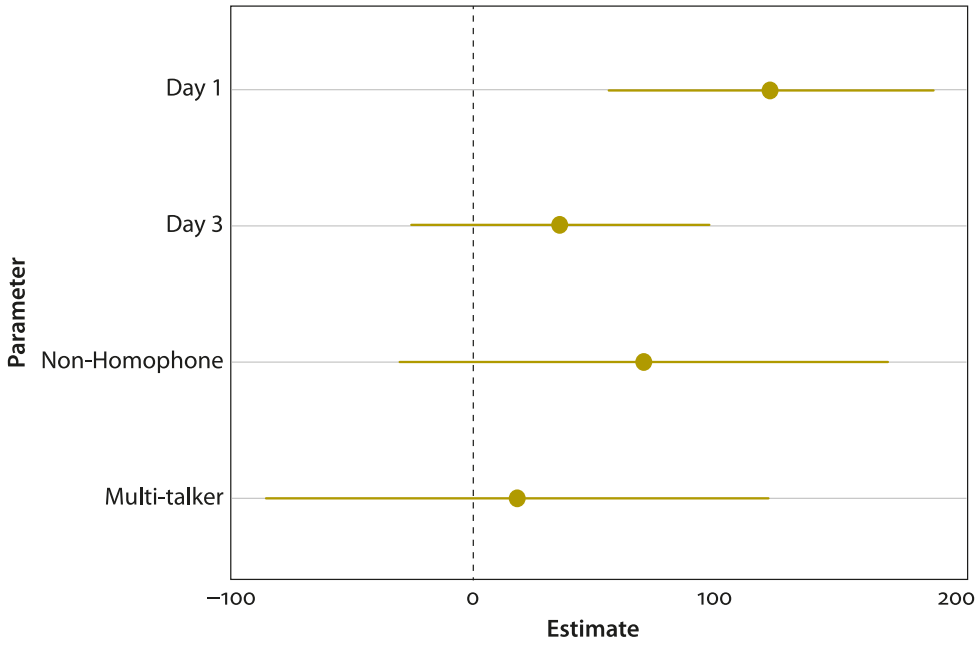

Figure 6. Estimates from mixed-effect linear regression model predicting correct RTs. Blue indicates a negative effect whereas yellow indicates a positive effect

Table 2. Summary of mixed-effects linear regression model predicting correct RTs

Parameter	Coefficient	SE	95% CI	$t(1166)$	p
(Intercept)	1762.99	54.50	[1656.18, 1869.81]	32.35	< .001
Day 1	122.99	34.59	[55.21, 190.78]	3.56	< .001
Day 3	35.38	31.48	[−26.32, 97.08]	1.12	0.261
Non-Homophone	70.44	51.79	[−31.06, 171.95]	1.36	0.174
Multi-talker	17.75	53.20	[−86.53, 122.03]	0.33	0.739
Random effects	**Variance**	**Std. Deviation**			
Participant	18790	137.08			
Item	9729	98.64			

learners can still take advantage of phonological overlap between known and new words when being trained and tested on unfamiliar talkers with variable F0 ranges.

Overall, participants identified new words more accurately and faster as a function of daily training. On Day 2, participants in both talker groups were more accurate and faster than they were on Day 1, irrespective of homophone status. No additional effects were found on Day 2 or any effects on Day 3. Thus, our

exploratory analysis revealed evidence of lexical learning independent of homophone status and talker group.

We therefore did *not* find a tonal homophone advantage in adult L2 learners of Mandarin when talker variability was introduced in the testing phase, irrespective of whether it was also introduced in the training phase. Unlike the participants in Liu and Wiener (2020), participants in both our single-talker and multi-talker groups did not identify new words with homophones already known to them more accurately and/or faster than new words without known homophones. This suggests that the tonal homophone advantage in L2 word learning observed by Liu and Wiener (2020) may have been partially driven by an exceptionally high level of talker familiarity, since that study used a single speaker for both training and testing. The talker normalization required in the present study meant that the L2 learners struggled to categorize relevant segmental and tonal information across different talkers. Indeed, for the participants in Liu and Wiener (2020), it was a relatively 'easy' task to learn and recognize 20 words spoken by a single talker. The highly consistent signal-to-representation match resulted in word recognition accuracy rates near ceiling on Day 2 (.89) and Day 3 (.92). In contrast, the single-talker (.84) and multi-talker (.82) groups in the present study were still making quite a few tonal errors on Day 3. An exploratory post-hoc ANOVA using the data from Liu and Wiener (2020) and the present study confirmed that the group trained and tested on the same talker was more accurate than our single-talker and multi-talker groups, $F(2,166) = 6.04$, $p = .003$, $\eta 2 = .05$. This reduction in accuracy is in line with recent L2 tone recognition results that showed high talker variability increases tonal competition and often yields inaccurate recognition of syllable+tone words (Wiener et al., 2021; Wiener & Lee, 2020) and that exposure to multiple talkers with different F0 ranges is difficult for L2 learners who struggle with talker normalization (e.g. Lee et al., 2009, 2013; Lee & Wiener, 2020). We note that any benefit of talker variability may simply be delayed for L2 Mandarin learners. Our three-day training paradigm may have been too short to see a robust talker variability effect. Yet, we corroborated work showing how overnight consolidation of tones in L2 learners (e.g., Qin & Zhang, 2019) can result in improved learning and recognition. Future work may explore a more extended length of exposure.

Our results are compatible with Jiang's model (2000, 2018) and demonstrate how sensitive the formal stage of lexical development can be in which form-meaning connections are built. Our results suggest that the difficulty involved in L2 word learning and recognition may largely result from the difficulty of matching a developing mental representation to variable speech input. In our study, participants tested on multiple speakers seemingly learned the new words similarly, irrespective of whether they heard a single talker or 16 talkers in the

training phase. However, both groups were less accurate than the group trained and tested on the same talker by Liu and Wiener (2020). The challenge apparently lies in matching the incoming signal to these nascent representations given unfamiliar talkers and the acoustic variability of their speech (e.g., Hardison, 2003; Sommers & Barcroft, 2011).

Our results are also in line with hybrid models of the lexicon that store both speaker-specific exemplars and abstract representations (e.g., Mitterer, Chen, & Zhou, 2011). Evidence for abstract representations was demonstrated by both the single-talker and multi-talker group as they performed equally despite the different training talker (e.g., McQueen, Cutler, & Norris, 2006). Evidence for talker-specific exemplars is still a viable explanation for Liu and Wiener's results (2020).

Finally, although we did not find a homophone advantage in our data, we re-examined Liu and Wiener's data to clarify whether L2 learners of Mandarin benefit from complete phonological overlap (as in the case of tonal homophones such as *shu1* 'book' and *shu1* 'uncle') or partial phonological overlap (as in the case of knowing many words with 'u' vowels and a high level tone like *qu1, du1, fu1*; see code online). We found that the observed homophone effect in Liu and Wiener can be explained by partial overlap of the onset-only and onset+tone, as well as by complete phonological overlap. Because the stimuli were not designed to test this specific hypothesis, we call for future studies to tease apart these variables, if possible.

We conclude by acknowledging the limitations to the study. Like Liu and Wiener (2020), the present study was unable to fully tease apart how tones, talkers, and homophones completely interact. Because the tested L2 learners knew a limited number of words, we were unable to test all potential tone pairs (e.g., *shu1* and *shu3* were tested but not *shu2-shu3, shu1-shu4*, etc.). Future studies with more advanced learners and with disyllabic nouns (e.g., Chang & Bowles, 2015) are needed. It is also unclear whether different phonological overlap in terms of a syllable, tone or vowel may affect L2 word learning. A design closer to Stamer and Vitevitch's (2012) in which each phonological category is systematically manipulated may yield different results. Finally, to what degree individual differences contributed to the present results (e.g., Bowles et al., 2016; Perrachione et al., 2011) remains unclear. We found null effects of Tonometric pitch perception abilities: our post-hoc exploratory correlation analyses revealed that pitch perception abilities were neither correlated with 4-AFC accuracy ($r = -.28$, $p = .09$) nor 4-AFC correct RT ($r = -.13$, $p = .43$). Other potential individual differences such as working memory (e.g., Juffs & Harrington, 2011) and inhibitory control (e.g., Linck, Hoshino, & Kroll, 2008) may be explored in future studies.

To conclude, we found that talker variability disrupts any homophone advantage that L2 Mandarin learners may be able to use. In combination with Liu and Wiener (2020), these findings suggest that specific phonological information (e.g., homophones) is stored in the lexicon. For beginner L2 Chinese learners, access to homophones is modulated by talker familiarity. Only when learners were exceptionally familiar with a single speaker's voice did they show a homophone advantage in new spoken word learning. Once multi-talker speech was used in the testing phase, regardless of whether participants were exposed to the multi-talker speech during the training, then a homophone advantage did not emerge in learning the new spoken words.

References

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition*, *27*, 387–414. https://doi.org/10.1017/S0272263105050175

Barcroft, J., & Sommers, M. S. (2014). Effects of variability in fundamental frequency on L2 vocabulary learning: A comparison between learners who do and do not speak a tone language. *Studies in Second Language Acquisition*, *36*, 423–449. https://doi.org/10.1017/S0272263113000582

Boersma, P. & Weenink, D. (2018). Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0. *43*.

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, *66*(4), 774–808. https://doi.org/10.1111/lang.12159

Bradlow, A. & Pisoni, D. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, *106*(1), 2074–2085. https://doi.org/10.1121/1.427952

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cueweighting and lexical tone learning. *Journal of the Acoustical Society of America*, *128*, 456–465. https://doi.org/10.1121/1.3445785

Chang, C. B., & Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *The Journal of the Acoustical Society of America*, *138*(6), 3703–3716. https://doi.org/10.1121/1.4937612

Chao, Y. R. (1965). *A grammar of spoken Chinese*. University of California Press.

Colantoni, L., Steele, J., Escudero, P., & Neyra, P. R. E. (2015). *Second Language Speech*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087636

Dong, H., Clayards, M., Brown, H., & Wonnacott, E. (2019). The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones. *PeerJ*, *7*, e7191. https://doi.org/10.7717/peerj.7191

Dong, Y., Gui, S., & MacWhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition*, *8*(3), 221–238. https://doi.org/10.1017/S1366728905002270

Duanmu, S. (2007). *The Phonology of Standard Chinese*. 2nd Ed. Oxford University Press.

Dumay, N., & Gaskell, M.G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*(2), 367–413. https://doi.org/10.1111/j.1467-9922.2010.00613.x

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, *130*(4), 206–212. https://doi.org/10.1121/1.3629144

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*, 233–277.

Hardison, D.M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*, 495–522. https://doi.org/10.1017/S0142716403000250

Hardison, D.M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, *26*(4), 579–596. https://doi.org/10.1017/S0142716405050319

Ho, A.T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, *33*, 353–367. https://doi.org/10.1159/000259792

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47–77. https://doi.org/10.1093/applin/21.1.47

Jiang, N. (2002). Form–meaning mapping in vocabulary acquisition in a second language *Studies in Second Language Acquisition*, *24*(4), 617–637. https://doi.org/10.1017/S0272263102004047

Jiang, N. (2013). *Conducting reaction time research in second language studies*. Routledge. https://doi.org/10.4324/9780203146255

Jiang, N. (2018). *Second Language Processing: An Introduction*. Routledge. https://doi.org/10.4324/9781315886336

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*(2), 137–166. https://doi.org/10.1017/S0261444810000509

Kroll, J.F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, *18*(2), 137–171. https://doi.org/10.1191/0267658302sr2010a

Kuhl, P.K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Leach, L. & Samuel, A. (2007). Lexical configuration and lexical engagement: When adults learn new words, *Cognitive Psychology*, *55*(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics*, *11*, 373–382. https://doi.org/10.1016/S0095-4470(19)30836-8

Lee, C.-Y., Tao, L., & Bond, Z.S. (2009). Speaker variability and context in the identification of fragmented Mandarin tones by native and non-native listeners. *Journal of Phonetics*, *37*, 1–15. https://doi.org/10.1016/j.wocn.2008.08.001

Lee, C.-Y., Tao, L., & Bond, Z.S. (2013). Effects of speaker variability and noise on Mandarin tone identification by native and non-native listeners. *Speech, Language and Hearing*, *16*(1), 1–9. https://doi.org/10.1179/2050571X12Z.0000000003

Lee, C.-Y., & Wiener, S. (2020). Acoustic-Based and Knowledge-Based Processing of Mandarin Tones by Native and Non-native Speakers. In: Liu, H., Tsao, F., Li, P. (eds) *Speech Perception, Production and Acquisition. Multidisciplinary approaches in Chinese languages* (pp. 37–57). Springer: Singapore. https://doi.org/10.1007/978-981-15-7606-5_3

Levelt, W.J. (1989). *Speaking: From intention to articulation* (Vol. *1*). MIT Press: Cambridge, MA. https://doi.org/10.7551/mitpress/6393.001.0001

Levelt, W.J. (1993). *Lexical access in speech production*. In Knowledge and language (pp. 241–251). Springer, Dordrecht.

Linck, J.A., Hoshino, N., & Kroll, J.F. (2008). Cross-language lexical processes and inhibitory control. *The Mental Lexicon*, *3*(3), 349–374. https://doi.org/10.1075/ml.3.3.06lin

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403–439). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-2037-8_16

Liu, J., & Wiener, S. (2020). Homophones facilitate lexical development in a second language. *System*, *91*, 102249. https://doi.org/10.1016/j.system.2020.102249

Lively, S.E., Logan, J.S., & Pisoni, D.B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the acoustical society of America*, *94*(3), 1242–1255. https://doi.org/10.1121/1.408177

Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36. https://doi.org/10.1097/00003446-199802000-00001

Mandell, J. (2018). Tonometric. http://jakemandell.com/

McQueen, J.M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126. https://doi.org/10.1207/s15516709cog0000_79

Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science*, *35*(1), 184–197. https://doi.org/10.1111/j.1551-6709.2010.01140.x

Moore, C.B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, *102*, 1864–1877. https://doi.org/10.1121/1.420092

Nelson, D.G.K., Hirsh-Pasek, K., Jusczyk, P.W., & Cassidy, K.W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*(1), 55–68. https://doi.org/10.1017/S030500090001343X

Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/BF03206860

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46. https://doi.org/10.1111/j.1467-9280.1994.tb00612.x

Packard, J. (2000). *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486821

Pelzl, E. (2019). What makes second language perception of Mandarin tones hard?: A non-technical review of evidence from psycholinguistic research. *Chinese as a Second Language. The Journal of the Chinese Language Teachers Association, USA*, *54*(1), 51–78. https://doi.org/10.1075/csl.18009.pel

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of The Acoustical Society of America*, *130*, 461–472. https://doi.org/10.1121/1.3593366

Qin, Z., & Zhang, C. (2019). The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PloS one*, *14*(12). https://doi.org/10.1371/journal.pone.0221498

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*, 1318. https://doi.org/10.3389/fpsyg.2014.01318

Sommers, M. S., & Barcroft, J. (2011). Indexical information, encoding difficulty, and second language vocabulary learning. *Applied Psycholinguistics*, *32*(2), 417–434. https://doi.org/10.1017/S0142716410000469

Sommers, M. S., & Barcroft, J. (2007). An integrated account of the effects of acoustic variability in first language and second language: Evidence from amplitude, fundamental frequency, and speaking rate variability. *Applied Psycholinguistics*, *28*, 231–249. https://doi.org/10.1017/S0142716407070129

Stamer, M. K., & Vitevitch, M. S. (2012). Phonological similarity influences word learning in adults learning Spanish as a foreign language. *Bilingualism: Language and Cognition*, *15*(3), 490–502. https://doi.org/10.1017/S1366728911000216

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, *44*(6), 1321–1337. https://doi.org/10.1044/1092-4388(2001/103)

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, *49*(6), 1175–1192. https://doi.org/10.1044/1092-4388(2006/085)

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition*, *28*(3), 387–422. https://doi.org/10.1017/S0272263106060177

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, *2*(1), 45–58. https://doi.org/10.1017/S1366728999000140

Tan, L. H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. *Reading and Writing*, *10*(3–5), 165–200. https://doi.org/10.1023/A:1008086231343

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735.

Vitevitch, M. S., & Storkel, H. L. (2013). Examining the acquisition of phonological word forms with computational experiments. *Language and Speech*, *56*(4), 493–527. https://doi.org/10.1177/0023830912460513

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*, *64*(2–3), 122–144. https://doi.org/10.1159/000107913

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America 106*, 3649–3658. https://doi.org/10.1121/1.428217

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*, 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Wiener, S., Ito, K., & Speer, S.R. (2018). Early L2 spoken word recognition combines input-based and knowledge-based processing. *Language and Speech*, *61*(4), 632–656. https://doi.org/10.1177/0023830918761762

Wiener, S., Ito, K., & Speer, S.R. (2021). Effects of multi-talker input and instructional method on the dimension-based statistical learning of syllable-tone combinations: An eye-tracking study. *Studies in Second Language Acquisition*, *43*(1), 155–180. https://doi.org/10.1017/S0272263120000418

Wiener, S., & Lee, C.Y. (2020). Multi-talker speech promotes greater knowledge-based spoken Mandarin word recognition in first and second language listeners. *Frontiers in Psychology*, *11*, 214. https://doi.org/10.3389/fpsyg.2020.00214

Wiener, S., Lee, C.Y., & Tao, L. (2019). Statistical regularities affect the perception of second language speech: Evidence from adult classroom learners of Mandarin Chinese. *Language Learning*, *69*(3), 527–558. https://doi.org/10.1111/lang.12342

Wiener, S., & Tokowicz, N. (2021). Language proficiency is only part of the story: Lexical access in heritage and non-heritage bilinguals. *Second Language Research*, *37*(4), 681–695. https://doi.org/10.1177/0267658319877666

Wilcox, A., & Medina, A. (2013). Effects of semantic and phonological clustering on L2 vocabulary acquisition among novice learners. *System*, *41*(4), 1056–1069. https://doi.org/10.1016/j.system.2013.10.012

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, *23*(1), 41–69. https://doi.org/10.1017/S0272263101001024

Wolter, B. (2006). Lexical Network Structures and L2 Vocabulary Acquisition: The Role of L1 Lexical/Conceptual Knowledge. *Applied Linguistics*, *27*(4), 741–747. https://doi.org/10.1093/applin/aml036

Wong, P.C.M., & Perrachione, T.K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565–585. https://doi.org/10.1017/S0142716407070312

Zareva, A. (2007). Structure of the second language mental lexicon: how does it compare to native speakers' lexical organization?. *Second Language Research*, *23*(2), 123–153. https://doi.org/10.1177/0267658307076543

Zhou, X. & Marslen-Wilson, W. (1995). Morphological Structure in the Chinese Mental Lexicon, *Language and Cognitive Processes*, *10*(6), 545–600. https://doi.org/10.1080/01690969508407114

Zhu, J., Chen, X., Chen, F., & Wiener, S. (2022). Individuals with Congenital Amusia show degraded speech perception but preserved statistical learning for tone languages. *Journal of Speech, Language, and Hearing Research*, *65* (1), 53–69. https://doi.org/10.1044/2021_JSLHR-21-00383

Address for correspondence

Jiang Liu
Department of Languages
Literatures and Cultures & Linguistics Program
University of South Carolina
709 Humanities Office Building
1620 College Street
Columbia, SC, 29208
USA

jiangliu@mailbox.sc.edu

Co-author information

Seth Wiener
Department of Modern Languages
Carnegie Mellon University

sethw1@cmu.edu

Publication history

Date received: 31 May 2020
Date accepted: 21 January 2022
Published online: 18 March 2022