LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Statistical Regularities Affect the Perception of Second Language Speech: Evidence From Adult Classroom Learners of Mandarin Chinese

Seth Wiener,[a] Chao-Yang Lee,[b] and Liang Tao[b]
[a]Carnegie Mellon University and [b]Ohio University

This study investigated how adult second language (L2) learners of Mandarin Chinese use knowledge of phonological and lexical statistical regularities when acoustic information is insufficient for word recognition. A gating task was used to test intermediate L2 learners at two time points across a semester of classroom learning. Native Mandarin speakers (tested once) served as a control group. Mixed-effects modeling revealed that upon hearing truncated speech, L2 learners, like native speakers, identified high token frequency syllable–tone combinations more accurately than low token frequency syllable–tone combinations. Error analysis of correct syllable/incorrect tone responses revealed that native speakers made specific probability-based errors. L2 learners primarily demonstrated more acoustic-based errors but exhibited a trend toward greater probability-based errors during the second test. These findings are interpreted in light of L2 speech learning models that emphasize a statistical learning mechanism.

Keywords speech perception; Mandarin Chinese; gating task; statistical learning; lexical tone

---

The authors thank Margaret Harrison and Sarah Greenlee for their assistance with data collection. Pavel Trofimovich, Kiwako Ito, and several anonymous reviewers provided thoughtful and critical comments that substantially improved earlier versions of this article.

This article has been awarded an Open Materials badge. All materials are publicly accessible via the Open Science Framework at https://osf.io/c4zn6. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

Correspondence concerning this article should be addressed to Seth Wiener, Department of Modern Languages, Carnegie Mellon University, 160 Baker Hall, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States. E-mail: sethw1@cmu.edu

DOI: 10.1111/lang.12342

Introduction

Speech perception and spoken word recognition can be characterized as incremental processes. Listeners perceive fine-grained acoustic–phonetic information and map this input to stored mental representations in real time (see Cutler, 2012; Dahan & Magnuson, 2006; Diehl, Lotto, & Holt, 2004; Samuel, 2011). When the speech signal is limited or ambiguous, listeners can improve upon the signal-to-representation match by drawing on their previous linguistic experience (e.g., Bybee, 2001, 2010; Clayards, Tanenhaus, Aslin, & Jacobs, 2008; McMurray, Tanenhaus, & Aslin, 2002). One form of linguistic experience is exposure to the statistical regularities inherent in speech. Users of language possess computational abilities to detect and learn from these regularities (Romberg & Saffran, 2010; Saffran, Aslin, & Newport, 1996). This statistical learning mechanism is central to first language (L1) acquisition as it promotes phonological, lexical, and syntactical learning and strengthens perceptual pathways (e.g., Kuhl, 2004; Newport, 2016).

In this study, we examined how this learning mechanism affects second language (L2) learners' speech perception processes. Exploring statistical learning in a L2 bears theoretical significance, since L2 learners differ from L1 learners in terms of developmental factors, which include age of acquisition, prior linguistic knowledge, and a perceptual system attuned to the L1 input (Cutler, 2012), and instructional factors, such as amount of input and explicit training (Ellis, 2002; Ellis, O'Donnell, & Römer, 2013, 2015). As a result, most adult L2 learners demonstrate poor perception of nonnative speech sounds, increased L2 lexical competition, and limited L2 word recognition (e.g., Iverson et al., 2003; Strange, 1995; Weber & Cutler, 2004).

Yet, despite the incongruent L1–L2 learning conditions and challenges stemming from L2 speech perception, L2 learners may still rely on an innate statistical learning mechanism to improve their recognition of speech. Lab-based evidence suggests that adults—like infants—are able to detect novel frequency distributions of acoustic–phonetic cues, such as consonants and vowels, as well as their co-occurrence as syllables and words (e.g., Imai, Walley, & Flege, 2005; MacKay, Meador, & Flege, 2001; Meador, Flege, & MacKay, 2000; Wanrooij, Escudero, & Raijmakers, 2013). These findings imply that for adult L2 learners, statistical information regarding the likelihood of the L2 input may help bootstrap poor acoustic–phonetic cue perception and improve global spoken word recognition.

To examine the dynamic tradeoff that occurs between learners' perception of acoustic–phonetic cues and their statistical knowledge about the target language, we used Mandarin Chinese as the target language. Because Mandarin

CV(C) syllables (where C and V designate consonants and vowels) can map directly to a morpheme or word, Mandarin allows for a relatively straightforward examination of how speech input affects lexical processing (Myers, 2006, 2010; Packard, 1999, 2000; Zhou & Marslen-Wilson, 1994, 1995). This mapping allows us to probe how knowledge of phonological (i.e., syllable, tone) and lexical (i.e., syllable–tone combination as a word) regularities interacts with classroom L2 learners' perception of Mandarin speech.

Background Literature

The L2LP Model of Learners' Speech Perception and Recognition

Early models of L2 speech perception, such as the Speech Learning model or SLM (Flege, 1995; Flege, Schirru, & MacKay, 2003) and the Perceptual Assimilation Model or PAM (Best, 1995) and its extensions PAM–L2 (Best & Tyler, 2007) and PAM–S (So & Best, 2010, 2014), were designed to capture the initial stage of L2 learning in which learners lacked sufficient lexical representations and had limited experience with the target language and its statistical regularities. While these models have considerably advanced our understanding of L2 perception and learning of nonnative speech sounds (for an overview, see Colantoni, Steele, & Escudero, 2015), they were not explicitly designed to test hypotheses involving statistical learning and its effect on L2 speech perception.

The Second Language Linguistic Perception model or L2LP (Escudero, 2005, 2009) provides a clear hypothesis regarding L2 statistical learning over the course of a learner's developmental trajectory. L2LP posits that L2 learners—like L1 learners (e.g., Maye, Werker, & Gerken, 2002)—possess a specific learning mechanism that allows them to extract relative frequency distributions from the speech signal. Whereas previous lab-based research has primarily explored L2 distributional learning of individual speech segments, such as nonnative vowels (e.g., Escudero, Benders, & Wanrooij, 2011; Escudero & Williams, 2014), the present study extended the general idea of distributions affecting perception to include segments, suprasegmentals, and their co-occurrences as words, focusing on dimension-based statistical learning (Idemaru & Holt, 2011, 2014; Wiener, Ito, & Speer, 2016, 2018).

The revised L2LP Model (van Leussen & Escudero, 2015) contains four levels of representation: acoustic, phonetic, phonemic, and lexical (Figure 1). Recognizing words is characterized as a four-step process through connected adjoining levels with recognition decided by the relative strength of these connections. The initial state of this model is a copy of a learner's L1 perceptual system; over the course of L2 learning, the strengths of these connections are altered. Crucial to the present study, experience with a language through

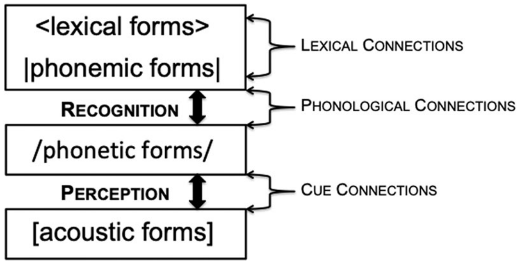

Figure 1 Interactive L2LP model (adapted from van Leussen & Escudero, 2015).

statistical learning can be conceptualized as the strengths of these connections. As an example, the L2LP model assumes that a beginner L1 English–L2 Mandarin learner relies on English categories and connections to perceive Mandarin speech. For many adult Mandarin learners with limited L2 classroom opportunities, these acoustic–phonetic cue connections remain weak for a considerable amount of time. Much of the early L2 tone perception literature demonstrates just how difficult acoustic–phonetic perception of tone is for many nonnative listeners (e.g., Hallé, Chang, & Best, 2004; Wang, Jongman, & Sereno, 2001; Wang, Spence, Jongman, & Sereno, 1999; Wayland & Guion, 2004).

As Mandarin classroom-based L2 learning shifts from pronouncing individual sounds and tones to producing word-, phrase-, and sentence-level communication, learners receive substantially more input. This includes tonal minimal pairs that require accurate perception of fine-grained acoustic–phonetic cues. The L2LP model posits that perceptual learning of these cues trickles down through "meaning-driven lexical learning" (van Leussen & Escudero, 2015, p. 3). Learners' perceptual pathways are updated over time as more frequent or more probable connections become stronger and less frequent, or less probable connections becoming weaker. In an interactive L2LP model, stronger lexical and phonological connections can influence weaker acoustic–phonetic cue connections (e.g., McClelland, Mirman, & Holt, 2006). Learning within the L2LP model is thus characterized as a probabilistic and meaning-driven process.

The present study is situated within an interactive L2LP framework to examine how distributional learning of lexical and phonological patterns affects learners' perception of acoustic–phonetic cues. We took a longitudinal approach to document how distributional learning of syllables, tones, and syllable–tone words occurs across multiple months within a L2 classroom. We demonstrate that while L2 Mandarin learners' acoustic–phonetic identification of tone and syllable–tone words may appear to plateau during classroom learning—as prior

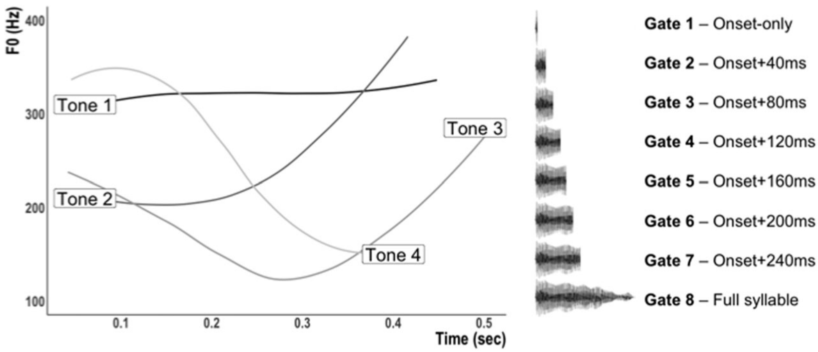

Figure 2 Four Mandarin tones spoken in isolation by a native female speaker and schematic illustration of the gating paradigm.

lab-based studies have typically documented (e.g., Hao, 2012)—L2 learners are accruing detailed knowledge of syllable and syllable–tone co-occurrence regularities. This distributional information is effectively extracted from the speech signal, even in structured L2 classroom contexts with relatively limited input.

L2 Perception of Mandarin Speech

Mandarin learners must perceive a speaker's fundamental frequency (F0) over time and categorize this information into one of four phonemic tones (Gandour, 1983; Howie, 1976). For example, as shown in Figure 2, the syllable *mai* can mean "to sell" when produced with a falling F0 (*mai4*) or "to buy" when produced with a low-dipping F0 (*mai3*). Even after several years of classroom learning, many L1 English–L2 Mandarin learners still struggle to distinguish between tones that share overlapping F0 starting points (Hao, 2012, 2018a). For instance, Figure 2 shows that tones 2 and 3 have a similar low onset F0, whereas tones 1 and 4 have a similar high onset F0 (Gottfried & Suiter, 1997; Shen & Lin, 1991). Previous L2 tone learning studies have explored how individual learner variability, different tone training paradigms, multispeaker input, and noise in the speech signal affect learners' perception of tone categories (e.g., Chandrasekaran, Sampath, & Wong, 2010; Leather, 1983; Lee, Tao, & Bond, 2009, 2010, 2013; Perrachione, Lee, Ha, & Wong, 2011; Wang et al., 1999; Wong & Perrachione, 2007). These studies have shown that while L2 learners' categorization of Mandarin tones improves as a result of short lab-based training and/or classroom L2 experience, learners do not typically reach nativelike perceptual abilities.

The present study expands prior tone-only categorization studies by requiring L2 learners to perceive consonant/vowel segmental information, suprasegmental tone information, and their combination as a syllable-tone word; that is, listeners must use all connections within the L2LP model. This allows for a test of whether statistical knowledge at the lexical (syllable–tone as a word) level and phonological (syllable or toneme) level (Ye & Connine, 1999) interacts with learners' acoustic–phonetic perception of speech.

Statistical Regularities in Mandarin Speech

Because nearly all syllable–tone co-occurrences map to an individual morpheme or word (Packard, 1999, 2000; Zhou & Marslen-Wilson, 1994, 1995), we examined syllable–tone words in this study. Although the majority of modern spoken Mandarin words are disyllabic by type, monosyllabic morphemes or words tend to be more frequent by token (Tao, 2015). For beginning L2 learners in particular, the majority of speech tokens follow a 1:1:1:1 mapping from syllable–tone to morpheme to word to written character (Myers, 2012). For the purpose of this study, a syllable–tone combination is hereafter referred to as a word, whereas a CV(C) segmental string, irrespective of tone, is hereafter referred to as a syllable.

The Mandarin syllable plays a crucial role in spoken word recognition and production (Chen, Chen, & Dell, 2002; You, Zhang, & Verdonschot, 2012). There are roughly 400 unique (C)V(C) Mandarin syllable types that vary in token frequency and homophone density (Duanmu, 2007, 2009; Packard, 1999, 2000; Wang, 1998). For example, the syllable *shi* occurs roughly 60 times more often than the syllable *zhou* in the spoken word corpus SUBTLEX–CH (Cai & Brysbaert, 2010).

At the word level, most syllable–tone combinations map to one or more homophonous words: *Zhou1* could refer to "week," "state/prefecture," "continent," "boat," and "porridge," among other semantically and orthographically distinct common forms. For the syllable *shi*, each *shi* + tone combination is associated with at least 10 homophonous words. As a result, the syllable *shi* has a homophone density nearly four times that of *zhou.* Over the course of acquisition, L2 learners may develop dense *shi* representations with many homophones but sparse *zhou* representations with relatively few homophones.

Lexical frequency can also affect the probability of a certain tone co-occurring with a syllable in speech. As an example, *zhou* has a high probability of appearing with the first tone as *zhou1* due to 周 "week" frequently occurring in speech. In contrast, *shi* has a high probability of appearing with the fourth tone as *shi4* due to its mapping to the copula 是 "to be." Given beginner L2 learners' relatively small lexicon and the skewed token frequency of "week" and "to be," learners may expect certain pitch patterns upon hearing the syllables *zhou* and *shi*. To summarize, during Mandarin spoken word recognition, listeners can rely on at least three forms of statistical information: syllable token frequency (e.g., *shi* occurs more frequently than *zhou*), syllable homophone density (e.g., *shi* has more homophones than *zhou*), and syllable-conditioned tonal probabilities (e.g., *shi4* and *zhou1* are the most likely *shi* + tone or *zhou* + tone words).

Evidence of native Mandarin listeners integrating statistical information with acoustic cues comes from Wiener and Ito's (2015) eye-tracking study. Participants were shown four frequency-controlled characters on screen, two of which shared the same syllable but differed in tone (e.g., *zhou1* and *zhou2*). Participants' looks to the four characters were recorded while listening to the target word. Upon hearing the initial segments in *zhou*, listeners first fixated on the character carrying the more probable tone—*zhou1*—before sufficient acoustic input disambiguated the target *zhou2*. As more F0 information unfolded, incorrect anticipatory looks to *zhou1* were rapidly revised. Importantly, a difference in predictive looks was observed between high and low token frequency syllables. Low token frequency syllables like *zhou* led to more probability-based processing. The authors argued that tone is less informative for high token frequency syllables because they typically co-occur with all tones to create dense homophone neighborhoods (e.g., Luce & Pisoni, 1998; Vitevitch & Luce, 1998, 1999; Vitevitch, Luce, Pisoni, & Auer, 1999). For low token frequency syllables with sparse homophone neighborhoods, tone is more informative; listeners have a much higher probability of hearing a *zhou1* word than any other *zhou* + tone word (Chen, Vaid, & Wu, 2009; Packard, 1999, 2000; Wang, Li, Ning, & Zhang, 2012).

In this study, we examined whether intermediate L2 listeners store and make use of the same phonological and lexical statistical information as observed in L1 listeners, and if so, how this information affects learners' perception of Mandarin syllables, tones, and their combination as words. In order to examine the underlying processes involved in speech perception, we made use of the gating paradigm.

Using Gating to Examine L2 Speech Perception

In gating studies, participants are asked to listen to increasingly longer fragments of a stimulus (e.g., phoneme, syllable, word, phrase) and then propose a linguistic unit based on the limited acoustic cues available (Cotton & Grosjean, 1984; Grosjean, 1980). With each increasing gate, more acoustic information is available to the listener (see Figure 2 for an illustration). As a result, gating identifies the amount of the acoustic signal needed to correctly identify the stimulus, that is, a stimulus' isolation point. Gating is a particularly valuable tool to investigate Mandarin spoken word identification because it allows for precise control of the amount of acoustic information given to listeners. Increasing longer stimuli provide an approximation of the isolation point of segmental information (individual phonemes and syllable), suprasegmental information (tone/toneme), as well as their combination as a syllable–tone word.

In a previous gating study, Wiener and Ito (2016) tested how much of the acoustic signal was needed for L1 Mandarin speakers to trigger learned statistical regularities and report expected syllables and tones in accordance with their prior linguistic experience. Participants heard either a high or low token frequency syllable with a tone that was either most or least probable for the syllable given the 33.5-million-word corpus SUBTLEX–CH (Cai & Brysbaert, 2010). After hearing only the onset and 40 milliseconds of the vowel (gate 2), native listeners correctly identified high token frequency (F+) syllables and their tones more accurately than low token frequency (F–) syllables and their tones. This syllable token frequency effect lasted until gate 5 (onset plus 160 milliseconds of the vowel).

Additionally, the authors' error analysis of correct syllable/incorrect tone responses revealed an effect of syllable-conditioned tonal probability. Errors of this nature suggested that participants heard sufficient acoustic information for segmental identification but insufficient F0 information for tone identification. In these responses, participants were forced to respond with either a statistically probable tone, given the segments and their knowledge of the language, or rely on limited F0 cues. These available F0 cues meant that tones 1 and 4 overlapped with a high onset F0 while tones 2 and 3 overlapped with a low onset F0 (see Figure 2). For example, upon hearing *zhou2* presented in an early gate, participants could have reported the more probable *zhou1* despite tone 1 and tone 2 having different F0 starting points. Alternatively, participants could have reported *zhou3* given that tones 2 and 3 start with similar low F0 points and tend to be acoustically similar for roughly the first 100 milliseconds (Moore & Jongman, 1997; Shen, Deutsch, & Rayner, 2013; Shen & Lin, 1991).

In the first three gates, during which the limited acoustic information available was sufficient for syllable identification but insufficient for tone identification, listeners reported the perceived syllable's most probable tone rather than an acoustically similar tone. This pattern of errors suggested that native listeners' reliance on tonal probability information could temporarily overshadow limited acoustic information. For example, upon hearing the onset and 40 milliseconds of truncated vowel information from *shi3*, participants often reported the more probable *shi4* morpheme rather than the acoustically similar *shi2*. The authors attributed these knowledge-based errors to listeners' greater experience with *shi4* morphemes, one of which is the Mandarin copula. This effect of tonal probability, however, was short-lived and statistically significant only for low token frequency (F–) syllables, presumably due to their sparse homophone density.

The Present Study

This study made use of the gating paradigm to examine to what degree the statistical information that naturally unfolds in an instructed L2 Mandarin classroom affects learners' identification of truncated speech. To accomplish this, we modified Wiener and Ito's (2016) gating stimuli. This ensured that the tested L2 learners were familiar with all potential items. We tested native listeners as a control group and L2 learners before and after 10–12 weeks of structured intermediate classroom learning. We examined whether L2 learners, like native speakers, are sensitive to syllable token frequencies and identify high token frequency syllables (F+) and their tones more accurately than low token frequency syllables (F–) and their tones. We also examined whether learners, like native listeners, are sensitive to syllable-conditioned tonal probabilities and report more probable syllable–tone combinations even in the face of incongruent acoustic information. Theoretically, these findings would serve as initial evidence that distributional learning of phonological (syllable) patterns and lexical (syllable–tone word) patterns can interact with L2 learners' acoustic–phonetic perception of segments and suprasegmentals.

Method

Participants

Fifteen Mandarin speakers from mainland China served as the native speaker group (8 female, 7 male; $M_{\text{age}} = 26.6$ years). All participants were international students studying at Ohio University and therefore spoke English as a L2. All participants self-reported being fluent in Mandarin only (i.e., no other local dialect), had normal speech and hearing, and were paid for their participation. Fifteen learners of L2 Mandarin served as the nonnative group (6 female, 9 male; $M_{\text{age}} = 19.7$ years). All participants spoke English as their L1, had normal speech and hearing, and were enrolled in an intermediate Mandarin class at Ohio University at the time of testing. This ensured that all participants were sufficiently familiar with *Pinyin*, the Mandarin romanization system used for keyboard input, and had adequate exposure to the language necessary to recognize the stimuli. On average, the L2 participants (henceforth, learners) had been studying for 2.4 years (three participants reported having limited exposure to Mandarin during high school). They were given class credit for fulfilling course assignments.

Materials and Procedure

Because the stimuli used in Wiener and Ito (2015, 2016) were deemed too difficult for L2 learners, a new set of 48 syllable–tone combinations were created (see Appendix S1 in the Supporting Information online). Items included 24 unique CV(N) syllables,[1] with N indicating an optional nasal coda. Two characteristics of the stimuli were manipulated: syllable token frequency and syllable-conditioned tonal probabilities. Twelve syllables were high token frequency (F+; $M_{\log} = 5.08$, *range* = 4.75–5.48), while the other 12 were low token frequency (F–; $M_{\log} = 3.84$, *range* = 3.18–4.35) according to the spoken word corpus SUBTLEX–CH (Cai & Brysbaert, 2010; see Wiener & Ito, 2015, for details regarding frequency calculations). This measure indicates the token frequency of a particular syllable, irrespective of the tone it co-occurs with. These condition means were statistically different from each other, $t(22) = 9.36, p < .001$.

For each of the 24 syllables, the most (P+) and least (P–) probable tones were selected based on learners' textbook vocabulary. Tonal probability therefore captures lexical information in the form of which tone is conditionally most likely to co-occur with the syllable (see Wiener & Ito, 2015, for details on probability calculations). While we cannot falsify whether learners had mental representations of all 48 tested words, we can surmise that they had previously encountered all 24 syllables with some tone prior to testing. For example, learners were familiar with *zhou1* as "week" and "state" as our P+ target but had most likely not learned a *zhou2* word (P–). We note that these probabilities are also in line with SUBTLEX–CH native speaker calculations.

While our original experimental design did not include tonal neighborhood density as an independent variable, we subsequently included calculations of homophone neighborhoods as a covariate in our analyses, as suggested by an external reviewer.[2] Homophone density was calculated using the number of unique head morphemes associated with each syllable in SUBTLEX–CH (*range* = 4–70). Given the need to use syllables L2 learners were familiar with, an unavoidable correlation was found between syllable token frequency and syllable homophone density ($r = .61, p < .01$). We acknowledge this potential confound and discuss its implications under Limitations.

A 28-year-old female from Beijing produced each of the 48 items. All items were recorded at 44,100 Hz in a sound-attenuated booth. Three additional native

Mandarin speakers (none of whom participated in the experiment) verified the syllable and the tone for each item with 100% agreement. A duration measurement confirmed previously reported temporal differences across tones (Howie, 1976; Moore & Jongman, 1997). Each of the 48 words was fragmented into eight gates using Praat (Boersma & Weenink, 2018). The first gate consisted of the syllable onset up to the beginning of the first regular periodic cycle of the vowel, in other words, with no F0 information. The remaining rime was separated into six 40-millisecond increments (2nd–7th gates; see Figure 1). The final (8th) gate contained the full syllable. This resulted in a total of 384 unique stimuli (48 words × 8 gates).

The stimuli were presented in a lab over headphones using Superlab 5 (Haxby, Parasuraman, Lalonde, & Abboud, 1993). Participants typed the *Pinyin* romanization of the perceived syllable and tone for each trial. Stimuli were duration blocked within each gate: After completing all 48 items in the first gate, participants began the 48 items in the second gate. Presentation order was pseudorandomized such that the same syllable did not appear in adjacent trials. The experiment lasted approximately 25 minutes.

Native speakers were tested only once. L2 learners were tested twice: once during the first week of their intermediate Mandarin course and again approximately 10–12 weeks later. During the first test, L2 learners were not made aware that they would be asked to return for a second test. Between tests, learners spent approximately 35 to 42 total hours in class. Each week of instruction included three classroom hours, 30 minutes in an oral practice lab, and quizzes on Monday (vocabulary and expressions) and Friday (grammar and general summary). Learners spent an additional estimated 40 hours outside of class self-studying, which included completing two written and two oral as signments per week. Learners also completed two midterm examinations, which included both oral and written tests. While learners were not explicitly taught any of the syllable–tone words used in the gating task as part of their assigned vocabulary between the two tests, learners were exposed to all the tested syllables with multiple different tones in speech, including syllable–tone homophones of the stimuli.

Results

All statistical analyses were carried out using the *lme4* package in R version 3.3.3 (Bates, Mächler, Bolker, & Walker, 2015). Inclusion of model variables, interactions, and random effects were determined using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017). Pairwise comparisons between L2 testing sessions were calculated using least square means from the *lsmeans*

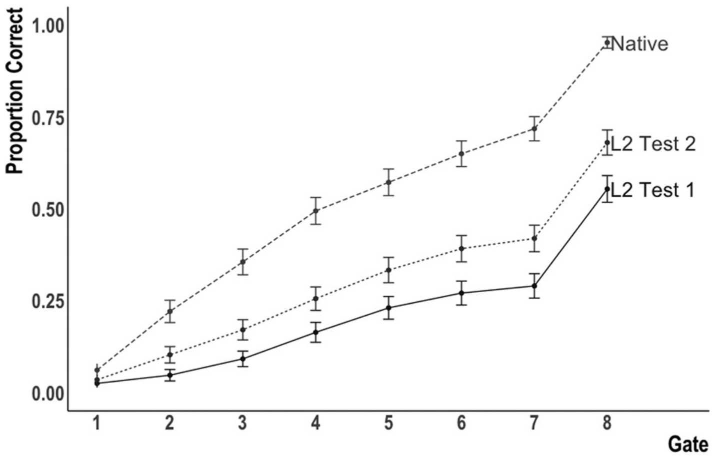

Figure 3 Mean correct syllable–tone accuracy by group and gate.

package (Lenth, 2016). Full output of all statistical models reported in this article, including model formulas and variable coding scheme, are available in Appendix S2 in the Supporting Information online.

Is Syllable–Tone Word Accuracy Affected by Distributional Information?
Figure 3 plots the correct syllable–tone word responses at each gate by listener group (error bars indicate 95% confidence intervals). At each gate, the same pattern was observed: Native speakers identified syllable–tone words most accurately, whereas L2 learners at Test 1 identified syllable–tone words least accurately. Of interest is the small, consistent accuracy improvement at each gate when the L2 group was tested roughly 10–12 weeks later. Mixed-effects logistic regression models revealed a main effect of group. The native group was more accurate than the L2 group at Test 1, $b = -2.32$, 95% CI [–1.67, –2.97], $SE = 0.33$, $z = -7.00$, $p < .001$, and Test 2, $b = -1.52$, 95% CI [–0.88, –2.18], $SE = 0.33$, $z = -4.63$, $p < .001$; the L2 group was more accurate at Test 2 compared to Test 1, $b = -0.79$, 95% CI [–1.44, –0.14], $SE = 0.33$, $z = -2.39$, $p = .02$. Main effects of syllable token frequency, $b = 0.97$, 95% CI [0.35, 1.60], $SE = 0.31$, $z = 3.08$, $p = .002$, and gate, $b = 0.66$, 95% CI [0.64, 0.69], $SE = 0.01$, $z = 54.90$, $p < .001$, were also found.

To further explore the between-group differences and identify the locus of the syllable token frequency effect, gate-by-gate analyses were carried out. Because accuracy was less than 4% at gate 1, these responses were removed from all subsequent analyses. The remaining responses were analyzed in four separate models to increase statistical power: gates 2–3, gates 4–5, gates 6–7, and gate 8. All four models revealed the same main effect of group. Native speakers were more accurate than the L2 group at Test 1 at gates 2–3, $b = -2.08$, 95% CI [–2.73, –1.42], $SE = 0.33$, $z = -6.19$, $p < .001$, gates 4–5, $b = -2.46$, 95% CI [–3.21, –1.71], $SE = 0.38$, $z = -6.42$, $p < .001$, gates 6–7, $b = -2.69$, 95% CI [–3.41, –1.98], $SE = 0.36$, $z = -7.38$, $p < .001$, and gate 8, $b = -3.33$, 95% CI [–4.27, –2.38], $SE = 0.48$, $z = -6.89$, $p < .001$. Similarly, native speakers were more accurate than the L2 group at Test 2 at gates 2–3, $b = -1.20$, 95% CI [–1.84, –0.56], $SE = 0.32$, $z = -3.68$, $p < .001$, gates 4–5, $b = -1.68$, 95% CI [–2.42, –0.93], $SE = 0.38$, $z = -4.41$, $p < .001$, gates 6–7, $b = -1.82$, 95% CI [–2.53, –1.11], $SE = 0.36$, $z = -5.04$, $p < .001$, and gate 8, $b = -2.59$, 95% CI [–3.55, –1.65], $SE = 0.48$, $z = -5.37$, $p < .001$. The L2 group was more accurate at Test 2 than at Test 1 in gates 2–3, $b = -0.85$, 95% CI [–1.12, –0.59], $SE = 0.13$, $z = -6.32$, $p < .001$, gates 4–5, $b = -0.68$, 95% CI [–0.87, –0.48], $SE = 0.09$, $z = -6.82$, $p < .001$, and gates 6–7, $b = -0.73$, 95% CI [–0.91, –0.56], $SE = 0.09$, $z = -8.05$, $p < .001$. At gate 8, the accuracy difference between the two tests was not significant ($p = .23$).

The main effect of syllable token frequency was found in gates 2–3 and gates 4–5. Subset analyses of the first two models confirmed that all listeners identified high token frequency (F+) syllables and their tones more accurately than low token frequency (F–) syllables and their tones: gates 2–3, $b = 1.88$, 95% CI [1.04, 2.72], $SE = 0.42$, $z = 4.41$, $p < .001$, and gates 4–5, $b = 1.52$, 95% CI [0.54, 2.49], $SE = 0.49$, $z = 3.06$, $p = .002$. Because the syllable token frequency effect could reflect facilitation on the syllable only or on the syllable and its tone, syllable error rates and tone error rates were further analyzed. In gates 2–3, each group's mean percentage of syllable errors (native = 53%; L2 Test 1 = 76%; L2 Test 2 = 69%) was similar to that of tone errors (native = 49%; L2 Test 1 = 72%; L2 Test 2 = 62%), $\chi^2(2) = 1.97$, $p = .37$. In gates 4–5, this pattern was repeated as listeners made a similar percentage of syllable errors (native = 33%; L2 Test 1 = 60%; L2 Test 2 = 50%) to tone errors (native = 26%; L2 Test 1 = 58%; L2 Test 2 = 47%), $\chi^2(2) = 1.05$, $p = .59$. Four models were built to test the effect of syllable token frequency on the log likelihood of syllable errors (in gates 2–3 and gates 4–5) and tone errors (in gates 2–3 and gates 4–5). Syllable token frequency had a significant effect on the likelihood of syllable errors in gates 2–3, $b = 2.68$, 95% CI [1.79, 3.56], $SE = 0.45$, $z = 5.92$, $p < .001$, and gates 4–5, $b = 2.11$, 95% CI [1.09, 3.13], $SE = 0.52$, $z = 4.06$, $p < .001$, while syllable token frequency had a null effect on the likelihood of tone errors ($p = .72$ in

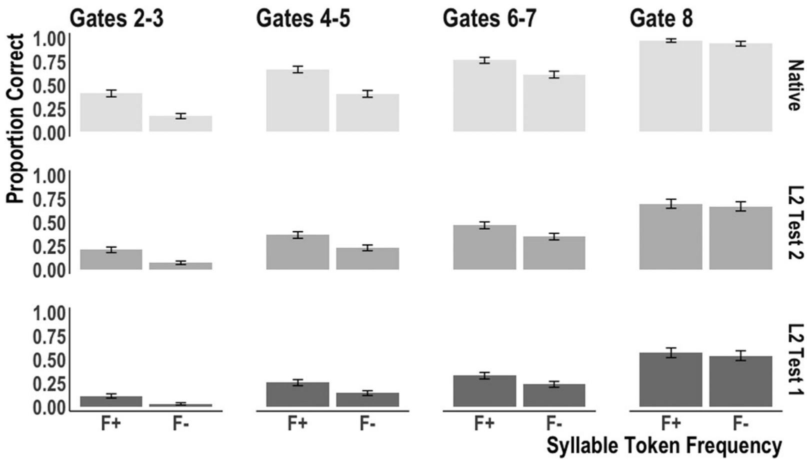

Figure 4 Mean correct word accuracy by syllable token frequency, group, and gate.

gates 2–3 and $p = .98$ in gates 4–5). Neither an effect of syllable homophone density nor an effect of tonal probability was found in any of the word accuracy models.

In summary, the syllable–tone word accuracy results indicate that from gate 2 until gate 7 (i.e., when the acoustic signal was truncated), the L2 group identified words more accurately at Test 2 than at Test 1 (Figure 3). No difference in word accuracy was observed at gate 8 between the two tests. From gate 2 (onset plus 40 milliseconds of the vowel) until gate 5 (onset plus 160 milliseconds of the vowel), native and L2 listeners at both tests identified high token frequency (F+) syllables (and their tones) more accurately than low token frequency (F–) syllables (and their tones). The lack of a two-way interaction between syllable token frequency and group indicated that L2 learners' distributional learning was consistent across tests; learners did not rely more on syllable token frequencies at Test 2 than at Test 1. Figure 4 shows that this pattern was consistent across groups.

Is Tone Accuracy Affected by Distributional Information?

Mean tone identification accuracy was calculated at gate 8, which corresponded to the full acoustic signal. This served as a measure of tone categorization accuracy (irrespective of the co-occurring segments) similar to previous L2 tone perception studies (e.g., Hao 2012, 2018a, 2018b; Wang et al., 1999). Because our stimuli did not control the sonority of the syllable onset, we do not examine tone accuracy at early gates or between tone type accuracy (cf. Wiener & Ito, 2016). Confidence intervals (95% CIs) revealed that native speakers were most accurate [0.99, 1.00], followed by L2 learners at Test 2 [0.79, 0.84], with L2 learners at Test 1 least accurate [0.71, 0.77].

Mixed-effects logistic regression models revealed that the native group identified tones more accurately than L2 learners at Test 1, $b = -5.51$, 95% CI [–7.51, –3.52], $SE = 1.02$, $z = -5.41$, $p < .001$, and at Test 2, $b = -4.56$, 95% CI [–6.55, –2.58], $SE = 1.01$, $z = -4.51$, $p < .001$. The L2 group did not differ in overall tone accuracy from Test 1 to Test 2 ($p = .38$). Neither a main effect of syllable frequency ($p = .50$) nor its interaction with group was found ($p = .71$). To summarize, intermediate L2 learners demonstrated small, yet statistically nonsignificant tone-only identification improvements at gate 8. These accuracy improvements were independent of any of the tested distributional information.

Are Tone Errors Affected by Distributional Information?

We first subset the data to examine responses in which participants correctly identified the syllable but incorrectly identified the tone. These errors suggest that participants heard sufficient acoustic information for syllable identification but insufficient F0 information for tone identification. These responses were next coded as either a probability-based error (e.g., *zhou1* response to *zhou2* stimulus because *zhou1* is more probable despite the dissimilar F0 onset) or an acoustic-based error (e.g., *zhou3* response to *zhou2* stimulus because the two tones have a similar F0 onset). Items in which the acoustically similar tone was also the most probable tone or instances in which the reported tone was neither more probable nor acoustically similar (e.g., reporting *zhou4* for *zhou2*, which occurred 20% of the time) were removed from subsequent analysis. In total, 1,416 responses were analyzed.

The empirical log of the ratio between probability-based errors and acoustic-based errors [log((probability error + 0.5) / (acoustic error + 0.5))] was calculated for each participant at each gate by syllable token frequency. A positive log ratio indicated greater probability-based errors while a negative log ratio indicated greater acoustic-based errors. Figure 5 shows the mean log ratio of the errors by group, gate, and syllable token frequency with error bars showing 1 standard error of the mean.

Figure 5 shows that for native speakers, with the exception of gates 2 and 3, the log ratio was roughly zero at all gates indicating an equal (and limited) number of both types of errors. In gates 2 and 3, a token frequency difference emerged. Native speakers made more acoustic-based errors for high token frequency (F+) syllables. In contrast, during the same two gates, native

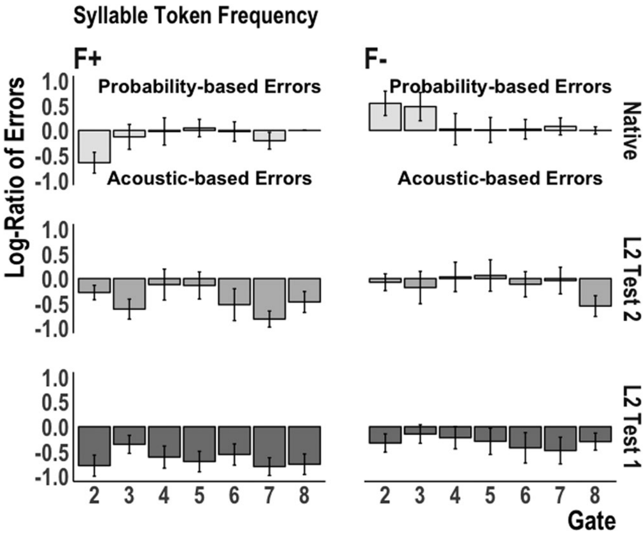

Figure 5 Mean log ratio of tone errors by syllable token frequency.

speakers made more probability-based errors for low token frequency (F–) syllables. For L2 learners at Test 1, both high and low token frequency syllables led to more acoustic-based errors across the gates overall, including late gates involving more of the acoustic signal. At Test 2, learners' log ratio increased and began approaching a nativelike pattern, particularly for the low token frequency (F–) syllables.

Following Wiener and Ito's (2016) statistical approach to increase power and roughly divide the observations into two equal windows, gates 2 and 3 were combined as an early window, while the remaining gates were combined as a late window. Separate analyses were run on the native and the L2 listeners, resulting in two mixed-effects linear regression models. Because syllable token frequency and syllable homophone density were correlated in our stimuli, syllable homophone density was dropped from the final model (see Zuur, Ieno, & Elphick, 2010, for multicollinearity discussions).

The native model revealed null effects of syllable frequency ($p = .61$) and window ($p = .25$). A statistically significant two-way interaction between window and syllable frequency was found, $b = -0.37$, 95% CI [–0.70, –0.13], $SE = 0.06$, $t = -6.55$, $p < .001$. Subset analyses confirmed that this interaction was restricted to the early window in which participants had a higher mean log ratio for low token frequency (F–) syllables than for high token frequency (F+)

syllables, $b = -0.18$, 95% CI [–0.33, –0.03], $SE = 0.07$, $t = -2.38$, $p = .02$. In the late window, the log ratio for high and low token frequency syllables did not differ ($p = .56$).

The L2 model revealed a main effect of syllable token frequency, $b = -0.15$, 95% CI [–0.32, –0.04], $SE = 0.04$, $t = -3.72$, $p < .001$. Subset analyses confirmed that the effect of syllable token frequency was observed in both tests: L2 learners had a higher mean log ratio for low frequency (F–) syllables than for high frequency (F+) syllables at Test 1, $b = -0.10$, 95% CI [–0.17, –0.03], $SE = 0.03$, $t = -2.96$, $p = .003$, and at Test 2, $b = -0.09$, 95% CI [–0.16, –0.01], $SE = 0.04$, $t = -2.31$, $p = .02$. There was a null effect of window ($p = .91$) while a marginal effect of test was found, $b = -0.22$, 95% CI [–0.30, 0.02], $SE = 0.13$, $t = -1.69$, $p = .10$. Similarly, a marginal two-way interaction between window and test was found, $b = 0.15$, 95% CI [0.07, 0.21], $SE = 0.08$, $t = 1.83$, $p = .06$. Subset analyses revealed that these marginal effects were the result of learners' increase from 92 total probability-based errors at Test 1, to 151 total probability-based errors at Test 2. A model testing this comparison confirmed that L2 learners had a higher mean log ratio in the late window at Test 2 as compared to Test 1, $b = -0.18$, 95% CI [–0.31, –0.05], $SE = 0.06$, $t = -2.84$, $p = .007$; no such difference was found in the early window across the tests ($p = .61$).

To summarize, native speakers made more acoustic-based errors in gates 2 and 3 for high token frequency (F+) syllables and more probability-based errors in gates 2 and 3 for low token frequency (F–) syllables. L2 learners primarily made acoustic-based errors at every gate across both tests. These acoustic-based errors, however, were more common for high token frequency (F+) syllables than for low token frequency (F–) syllables. At Test 2, L2 learners showed a marginal tendency to make more probability-based errors in later gates as compared to the same gates at Test 1.

Discussion

The Dynamic Tradeoff Between Statistical Information and Acoustic–Phonetic Cues

We tested the hypothesis that knowledge of Mandarin syllable token frequency and syllable–tone co-occurrence probabilities affect the perception of Mandarin syllables, tones, and their combination as words. We used the gating paradigm to test native Mandarin speakers once, and L2 learners at the start and end of a semester of intermediate classroom learning. We reported three findings. First, we found that at gate 8, when the speech signal was full, learners demonstrated small accuracy improvements in syllable–tone word identification and tone-only categorization (Figure 3). These changes in accuracy between Test 1 and Test 2, however, were not statistically significant. On the surface it appeared that adult L2 Mandarin learners had made little progress in their tone perception and spoken word recognition despite 10–12 weeks of additional explicit classroom learning. This finding corroborates prior evidence that beginner and intermediate L2 learners' perception and production of tone is often resistant to improvement and much less accurate than that of native listeners (e.g., Hao, 2012, 2018a, 2018b; Shen, 1989; Wiener, 2017; Yang, 2016; Yang & Chan, 2010).

However, despite the apparent learning plateau observed at gate 8, our second finding revealed that L2 learners undeniably improved in their perceptual abilities. Analyses of gates 2 through 7 confirmed that when the speech signal was truncated, greater experience with the language resulted in earlier and higher accuracy of syllable–tone word identification (Figure 3). Additionally, in gates 2 through 5 (i.e., onset plus 40 milliseconds to 160 milliseconds of the vowel), high token frequency (F+) syllables and their tones were identified more accurately than low token frequency (F–) syllables and their tones. This pattern was observed in native speakers and L2 learners at each test (Figure 4). Post hoc analyses revealed that this token frequency effect facilitated listeners' syllable-only identification, which led to improved syllable–tone word recognition. While ample research has demonstrated this robust syllable token frequency effect in native Mandarin speakers (e.g., Wang et al., 2012; Wiener & Turnbull, 2016; Zhou & Marslen-Wilson, 1994, 1995), to our knowledge, this is the first reported evidence of L2 classroom learners' sensitivity to Mandarin phonological statistical regularities. Despite limited exposure to Mandarin in a non-immersion linguistic environment, L2 learners used their knowledge of CV(C) token frequency to better identify words from truncated speech. Remarkably, L2 learners' knowledge of frequent phonological patterns was used to bootstrap perception even at Test 1. These results suggest that prior to testing (i.e., during the first year of classroom learning), L2 learners began encoding statistical and phonological information in L2 representations (e.g., Hayes-Harb, 2007; Hayes-Harb & Masuda, 2008).

Our third finding revealed important differences between native and nonnative listeners in early gates. Analyses of correct syllable/incorrect tone responses (Figure 5) revealed that for native listeners, specific lexical knowledge resulted in probability-based tone errors. When presented with the onset and up to 80 milliseconds of the vowel (i.e., gates 2 and 3), native listeners more often responded with an acoustic-based error for a high token frequency (F+) syllable and with a probability-based error for a low token frequency (F–)

syllable, thus corroborating Wiener and Ito's results (2016). However, a difference was observed between the results of these two studies in the first two gates for high-frequency (F+) syllables. In the present study, native listeners made more acoustic-based errors on these F+ syllables. We attribute this difference to the present study's stimuli, which included more high-frequency sonorant-initial syllables than Wiener and Ito's stimuli.

For L2 learners, tone errors were primarily acoustic-based across both tests. Learners also made more acoustic-based errors for high token frequency (F+) syllables than they did for low token frequency (F–) syllables, confirming learners' reliance on syllable token frequency information. Unlike the native listeners tested, L2 listeners continued to make tone errors in the later gates, including the last two gates. For L2 learners, a marginal two-way interaction between window and test was found: In the last four gates, L2 learners made over 60% more probability-based errors at Test 2. Therefore, learners showed an emerging tendency to rely more heavily on their prior lexical experience than on the truncated acoustic–phonetic cues from the speech signal.

Implications for L2 Speech Perception

The revised L2LP Model (van Leussen & Escudero, 2015), illustrated in Figure 1, assumes that when L1 English listeners first step into the L2 Mandarin classroom, they use their English perceptual system to perceive Mandarin speech. Acoustic, phonetic, phonemic, and lexical connections are, therefore, all relatively weak when an adult learner begins acquiring Mandarin. Over time, these connection strengths change as a function of the input. For the L2 learners tested in the present study, neither tone-only categorization accuracy nor word identification accuracy at the final gate improved between tests. On the surface, it therefore appeared that connection strengths did not change despite the intervening 10–12 weeks of classroom input.

Yet, when faced with truncated speech, learners demonstrated measurable sensitivity to phonological distributional regularities. Stronger phonological connections interacted with weaker acoustic–phonetic cue connections to improve the identification of high token frequency syllable–tone words. As an example, L2 learners were exposed to a variety of words containing the frequent syllable *da* during their classroom learning, which strengthened learners' acoustic–phonetic–phonemic connections for *da* syllable/toneme representations. As a result, L2 learners improved their mean identification of *da* items across all gates by nearly 16% from Test 1 to Test 2. This is most apparent at early gates (Figure 3) in which high token frequency syllable representations like *da* were more likely to be activated when listening to ambiguous or truncated speech (e.g., Bybee, 2001, 2010; Bybee & Hopper, 2001; Grosjean, 1980). In contrast, representations like *gua*—a low token frequency syllable—were less likely to be activated in early gates. Given the relatively limited number of *gua* exemplars learners were exposed to in class, acoustic–phonetic–phonemic connections remained weak for *gua* syllables, tones, and words (e.g., Goldinger, 1998). Learners demonstrated only a 5% improvement in mean identification of *gua* items across all gates from Test 1 to Test 2.

Additionally, our results demonstrated a trend toward stronger L2 lexical connections interacting with learners' acoustic–phonetic cue connections. Yet, the timing and degree to which lexical information influenced participants' perception differed between the native and nonnative speakers. For native speakers, probability-based tone errors were observed primarily on low token frequency (F–) syllables. For instance, lexical knowledge that *zhou* most often occurs as "week" or "state" was used to predict *zhou1* upon hearing the onset and up to 80 milliseconds of the vowel. This occurred in native listeners, despite participants hearing limited *zhou2* input (i.e., incongruent F0 information). In contrast, intermediate L2 learners demonstrated only a marginal increase in probability-based errors at Test 2 compared to Test 1. L2 learners also required more of the acoustic signal in order to trigger knowledge-based processing. At Test 2, learners' probability-based errors were primarily made after the onset and 120 milliseconds of the vowel.

We attribute this lack of nativelike lexical sensitivity to L2 learners' relatively small lexicon, restricted distributional knowledge, and need for greater acoustic information to initiate lexical access. Unlike advanced Mandarin–English bilinguals who recognize spoken English words with the same amount of acoustic information as monolinguals (Li, 1996), the intermediate L2 Mandarin learners tested clearly required more acoustic information to initiate lexical access. Because more acoustic information was needed, distributional knowledge of likely words was integrated later as a result. Learners' available syllable–tone distributional information may have also been restricted to a small subset of the tested syllables. While the 10–12 weeks of input between Test 1 and Test 2 included exposure to hundreds of new morphemes/words, what percentage of that input was homophonous with learners' previous representations remains an empirical question. For instance, learners were undoubtedly exposed to *da4* morphemes or words more often than *da2* morphemes or words, since the former occurs in the frequent compound *da4xue2* "university" and *da4* alone can mean "big." *Da4* lexical connections were presumably more active during classroom learning than *da2* lexical connections and resulted in a well-defined *da* + tone distribution. It is unclear whether learners were exposed to similar distributions for all the syllables tested.

Our view of speech learning assumes that with enough exposure, more advanced L2 learners should approach nativelike perception—a possible end state in the L2LP model. Advanced learners should also approach nativelike timing with respect to integrating lexical distributional information with acoustic–phonetic cues. For a L2 classroom learner without a lexicon containing a sufficiently large number of tonal minimal pairs and corresponding distributional information (such as the tested participants), acoustic–phonetic cue connections may remain comparatively weak for an extended period of time. We argue that a greater number of tonal minimal pairs in the lexicon necessitates finer perception of acoustic–phonetic cues through "meaning-driven lexical learning" (van Leussen & Escudero, 2015, p. 3), which ultimately results in nativelike tone perception. From this perspective, L2 tone learning should plateau during classroom acquisition (e.g., Hao, 2012; Wiener, 2017). Our results indicate either a longer window of L2 learning and/or input containing more tonal minimal pairs (and their distributional information) may be required in order to produce more robust changes in learners' behavior so they may ultimately reach nativelike abilities.

Limitations and Future Research

The main limitation of this study is that the stimuli did not adequately allow for a test of whether the reported tonal probability effect was driven by syllable token frequency, syllable homophone density, and/or their interaction. Given the need to use syllables familiar to our L2 learners, we were unable to use roughly a third of the Mandarin syllabary as items. If syllable homophone density affects tone perception as speculated, then L2 learners would need to acquire a sufficient number of homophones across different syllable types in order for an effect to emerge. This would require testing advanced L2 learners, since undoubtedly the stimuli would include low token frequency syllables with sparse homophone densities; in other words, relatively rare syllables associated with infrequent morphemes or words. While there remains a dearth of studies on advanced L2 Mandarin learners' tone perception, recent findings by Pelzl, Lau, Guo, and DeKeyser (2018) suggest that learners with an average of 10 years of L2 study approach nativelike tone-only categorization and syllable–tone lexical decision accuracy. Advanced L2 learners may therefore have a sufficiently sized lexicon and demonstrate nativelike sensitivity to phonological and lexical distribution information as we predict.

Similarly, it is unclear to what degree L2 learners' knowledge of syllable token frequency is correlated with their knowledge of lexical frequency. Because our response method did not require participants to write or select a specific Chinese character—keyboard input does not explicitly require tone identification and would have biased participants to pick the most frequent character—it is unknown whether participants' responses always captured their knowledge of the most frequent word for the reported syllable–tone combination. Future studies will need to tease apart these two variables.

As with any experimental study, the present results reflect the relatively limited random sample of native and L2 learners tested. Given the inherent variability associated with language learning and perception (e.g., Bradlow & Pisoni, 1999), attempts were made to increase statistical power whenever possible. Observed power simulations were carried out on our mixed-effects models using the *simr* package in R (Green & MacLeod, 2016). For the main effect of syllable token frequency in gates 2–3 and gates 4–5, the 95% CI of the observed power was [0.7, 1.0] for both models. For the error analysis models, the two-way interaction observed in the native speaker model had a 95% CI of [0.6, 1.0], while the main effect of syllable token frequency for the L2 learners had a 95% CI of [0.7, 1.0].

Conclusion

This study demonstrated that intermediate L2 learners' perception of Mandarin speech changes as a result of experience with phonological and lexical statistical regularities in the input. Higher level linguistic strategies derived from learners' knowledge of the language's statistical regularities appear to play as important a role in L2 Mandarin spoken word recognition as lower level detection of highly variable acoustic–phonetic cues. Interactive L2 speech learning models that implement linguistic knowledge as strengths between acoustic, phonetic, phonemic, and lexical connections (e.g., L2LP) may model these findings. Future research will need to examine whether additional exposure to the full distribution of co-occurring speech cues, a larger longitudinal window of L2 learning, and a higher level of L2 proficiency affect learners' integration of statistical information with acoustic–phonetic cues.

Final revised version accepted 6 January 2019

Notes

1 Given the constraints on our stimuli, we were unable to control the sonority of the syllable onsets in this study.

2 We thank the anonymous reviewer for bringing this to our attention.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer (Version 6.0.37) [Computer software]. http://www.fon.hum.uva.nl/praat

Bradlow, A. R., & Pisoni, D. B. (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*, 2074–2085. https://doi.org/10.1121/1.427952

Bybee, J. L. (2001). *Phonology and language use*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511612886

Bybee, J. L. (2010). *Language, usage and cognition*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Bybee, J. L., & Hopper, P. J. (2001). *Frequency and the emergence of linguistic structure* (Vol. *45*). Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/tsl.45

Cai, Q., & Brysbaert, M. (2010). SUBTLEX–CH: Chinese word and character frequencies based on film subtitles. *Plos ONE*, *5*, e10729. https://doi.org/10.1371/journal.pone.0010729

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

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*, 751–781. https://doi.org/10.1006/jmla.2001.2825

Chen, H. C., Vaid, J., & Wu, J. T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language and Cognitive Processes*, *24*, 967–982. https://doi.org/10.1080/01690960902804515

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*, 804–809. https://doi.org/10.1016/j.cognition.2008.04.004

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech*. Cambridge, UK: Cambridge University Press.

Cotton, S., & Grosjean, F. (1984). The gating paradigm: A comparison of successive and individual presentation formats. *Perception & Psychophysics*, *35*, 41–48. https://doi.org/10.3758/BF03205923

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Boston, MA: MIT Press.

Dahan, D., & Magnuson, J. S. (2006). Spoken word recognition. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 249–283). New York, NY: Academic Press. https://doi.org/10.1016/B978-012369374-7/50009-2

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–179. https://doi.org/10.1146/annurev.psych.55.090902.142028

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). New York, NY: Oxford University Press.

Duanmu, S. (2009). *Syllable structure: The limits of variation*. New York, NY: Oxford University Press.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning*, *63*, 25–51. https://doi.org/10.1111/j.1467-9922.2012.00736.x

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2015). Usage-based language learning. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (163–180). New York, NY: Wiley. https://doi.org/10.1002/9781118346136.ch7

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. (Unpublished doctoral thesis). LOT Dissertation Series 113: Utrecht University.

Escudero, P. (2009). Linguistic perception of "similar" L2 sounds. *Phonology in Perception*, *15*, 152–190.

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*, 206–212. https://doi.org/10.1121/1.3629144

Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition*, *133*, 408–413. https://doi.org/10.1016/j.cognition.2014.07.002

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*, 467–491. https://doi.org/10.1016/S0167-6393(02)00128-0

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149–176.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279. https://doi.org/10.1037//0033-295x.105.2.251

Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, *25*, 207–231. https://doi.org/10.1006/jpho.1997.0042

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498. https://doi.org/10.1111/2041-210X.12504

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, *28*, 267–283. https://doi.org/10.3758/BF03204386

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

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*, 269–279. https://doi.org/10.1016/j.wocn.2011.11.001

Hao, Y. C. (2018a). Second language perception of Mandarin vowels and tones. *Language and Speech*, *61*, 135–152. https://doi.org/10.1177/0023830917717759

Hao, Y. C. (2018b). Contextual effect in second language perception and production of Mandarin tones. *Speech Communication*, *97*, 32–42. https://doi.org/10.1016/j.specom.2017.12.015

Haxby, J. V., Parasuraman, R., Lalonde, F., & Abboud, H. (1993). SuperLab: General-purpose Macintosh software for human experimental psychology and psychological testing. *Behavior Research Methods, Instruments, & Computers*, *25*, 400–405. https://doi.org/10.3758/bf03204531

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*, 65–94. https://doi.org/10.1177/0267658307071601

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*, 5–33. https://doi.org/10.1177/0267658307082980

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge, UK: Cambridge University Press.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1939–1956. https://doi.org/10.1037/a0025641

Idemaru, K., & Holt, L. L. (2014). Specificity of dimension-based statistical learning in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1009–1021. https://doi.org/10.1037/a0035269

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *The Journal of the Acoustical Society of America*, *117*, 896–907. https://doi.org/10.1121/1.1823291

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics*, *11*, 373–382.

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. (2010). Identification of multi-speaker Mandarin tones in noise by native and non-native listeners. *Speech Communication*, *52*, 900–910. https://doi.org/10.1016/j.specom.2010.01.004

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*, 1–9. https://doi.org/10.1179/2050571×12Z.0000000003

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*, 1–33. https://doi.org/10.18637/jss.v069.i01

Li, P. (1996). Spoken word recognition of code-switched words by Chinese–English bilinguals. *Journal of Memory and Language*, *35*, 757–774. https://doi.org/10.1006/jmla.1996.0039

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

MacKay, I. R., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica*, *58*, 103–125. https://doi.org/10.1159/000028490

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111. https://doi.org/10.1016/S0010-0277(01)00157-3

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*, 363–369. https://doi.org/10.1016/j.tcs.2006.06.007

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*, B33–B42. https://doi.org/10.1016/S0010-0277(02)00157-9

Meador, D., Flege, J. E., & MacKay, I. R. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition*, *3*, 55–67. https://doi.org/10.1017/s1366728900000134

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

Myers, J. (2006). Processing Chinese compounds: A survey of the literature. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 169–196). Oxford, UK: Oxford University Press.

Myers, J. (2010). Chinese as a natural experiment. *The Mental Lexicon*, *5*, 423–437. https://doi.org/10.1075/ml.5.3.09mye

Myers, J. (2012). Chinese as a natural experiment. In G. Libben, G. Jarema, & C. Westbury (Eds.), *Methodological and analytic frontiers in lexical research* (pp. 155–169). Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/bct.47.09mye

Newport, E. L. (2016). Statistical language learning: Computational, maturational, and linguistic constraints. *Language and Cognition*, *8*, 447–461. https://doi.org/10.1017/langcog.2016.20

Packard, J. L. (1999). Lexical access in Chinese speech comprehension and production. *Brain and Language*, *68*, 89–94. https://doi.org/10.1006/brln.1999.2102

Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511486821

Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2018). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition*. Published online 15 May 2018. https://doi.org/10.1017/S0272263117000444

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (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

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 906–914. https://doi.org/10.1002/wcs.78

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology*, *62*, 49–72. https://doi.org/10.1146/annurev.psych.121208.131643

Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of Chinese Language Teachers Association*, *24*, 27–47.

Shen, J., Deutsch, D., & Rayner, K. (2013). On-line perception of Mandarin Tones 2 and 3: Evidence from eye movements. *The Journal of the Acoustical Society of America*, *133*, 3016–3029. https://doi.org/10.1121/1.4795775

Shen, X. S., & Lin, M. (1991). The use of prosody in disambiguation in Mandarin. *Language and Speech*, *34*, 145–156. https://doi.org/10.1177/002383099103400202

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*, 273–293. https://doi.org/10.1177/0023830909357156

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*, 195–221. https://doi.org/10.1017/S0272263114000047

Strange, W. (Ed.). (1995). *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore, MD: York Press.

Tao, H. (2015). Profiling the Mandarin spoken vocabulary based on corpora. In W. S.-Y. Wang & C. Sun (Eds.), *The Oxford handbook of Chinese linguistics* (pp. 336–347). Oxford, UK: Oxford University Press.

van Leussen, J. W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*, 1000. https://doi.org/10.3389/fpsyg.2015.01000

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*, 325–329. https://doi.org/10.1111/1467-9280.00064

Vitevitch, M. S., & Luce, P.A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374–408. https://doi.org/10.1006/jmla.1998.2618

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, *68*, 306–311. https://doi.org/10.1006/brln.1999.2116

Wang, H. S. (1998). An experimental study on the phonetic constraints of Mandarin Chinese. In B. K. Tsou (Ed.), *Studia linguistica serica* (pp. 259–268). Hong Kong: City University of Hong Kong Language Information Sciences Research Center.

Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language*, *78*, 332–348. https://doi.org/10.1006/brln.2001.2474

Wang, W., Li, X., Ning, N., & Zhang, J. X. (2012). The nature of the homophone density effect: An ERP study with Chinese spoken monosyllable homophones. *Neuroscience Letters*, *516*, 67–71. https://doi.org/10.1016/j.neulet.2012.03.059

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

Wanrooij, K., Escudero, P., & Raijmakers, M. E. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, *41*, 307–319. https://doi.org/10.1016/j.wocn.2013.03.005

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*, 681–712. https://doi.org/10.1111/j.1467-9922.2004.00283.x

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. (2017). Changes in early L2 cue-weighting of non-native speech: Evidence from learners of Mandarin Chinese. In F. Lacerda (Ed.), *Proceedings of Interspeech 2017* (pp. 1765–1769). Stockholm, Sweden: Interspeech 2017. https://doi.org/10.21437/Interspeech.2017-289

Wiener, S., & Ito, K. (2015). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition & Neuroscience*, *30*, 1048–1060. https://doi.org/10.1080/23273798.2014.946934

Wiener, S., & Ito, K. (2016). Impoverished acoustic input triggers probability-based tone processing in mono-dialectal Mandarin listeners. *Journal of Phonetics*, *56*, 38–51. https://doi.org/10.1016/j.wocn.2016.02.001

Wiener, S., Ito, K., & Speer, S. R. (2016). Individual variability in the distributional learning of L2 lexical tone. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Speech Prosody 2016* (pp. 538–542). Boston, MA: Speech Prosody 2016. https://doi.org/10.21437/SpeechProsody.2016-110

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

Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech*, *59*, 59–82. https://doi.org/10.1177/0023830915578000

Wong, P. C., & 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

Yang, C. (2016). *The acquisition of L2 Mandarin prosody: From experimental studies to pedagogical practice*. Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/bpa.1

Yang, C., & Chan, M. K. (2010). The perception of Mandarin Chinese tones and intonation by American learners. *Journal of Chinese Language Teachers Association*, *45*, 7–36.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, *14*, 609–630. https://doi.org/10.1080/016909699386202

You, W., Zhang, Q., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PloS One*, *7*, e46595. https://doi.org/10.1371/journal.pone.0046595

Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processes*, *9*, 393–422. https://doi.org/10.1080/01690969408402125

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

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*, 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Syllable–Tone Combinations.
Appendix S2. Output of Statistical Models.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Do Classroom Learners of Mandarin Chinese Improve in Their Speech Perception?

What This Research Was About and Why It Is Important

For many native English speakers, Mandarin Chinese is often considered one of the more difficult foreign languages to learn. Adult learners regularly struggle to accurately understand Mandarin speech, which includes both sounds and tones that might not be familiar to speakers of English. As a result, learners tend to stop learning the language before reaching a high proficiency level. In this study, the researchers examined adult learners' perception of spoken Mandarin words at the start and end of a 15-week semester of intermediate Mandarin classroom learning. The researchers found that, when listening to a word, learners demonstrated little to no improvement in their perception accuracy for spoken Mandarin between the two tests. This finding supports previous research and learners' intuition that Mandarin classroom learning results in limited improvement. However, when listening to speech fragments (incomplete words containing only several initial sounds from a word), learners showed a large improvement between the two tests, particularly for frequent speech sounds that learners had heard many times in class. The researchers argued that the learners' perception of Mandarin speech improved but at a level not typically measured in language research.

What the Researchers Did

- The researchers recruited 15 native Mandarin speakers from China to establish how native listeners perform. These listeners were only tested once.
- The researchers recruited 15 native English speakers enrolled in an intermediate Mandarin language class at a university to examine how learners perform. These listeners were tested twice, roughly 10–12 weeks apart.
- The perception test involved learners hearing fragments of Mandarin words until the entire word was presented. The learners first heard short word fragments over headphones and were asked to type (that is, identify) the word they believed they heard using a keyboard. For example, participants heard only the "d" consonant sound of the word *da4* meaning "big." Gradually they heard increasingly longer portions of the word until they heard the full word.
- The learners heard a variety of Mandarin words, some of which learners frequently heard in class while other words were rarely heard in class.

What the Researchers Found

- The native Mandarin speakers were able to correctly identify words—both from word fragments and when the entire word was heard—more accurately than the classroom learners.
- The native speakers and the learners identified frequent words more accurately than rare words.
- When listening to word fragments, both the native speakers and the learners likely relied on their experience with the language, which for learners amounted to hearing these words spoken in their classes, to guess what the word might be.

Things to Consider

- The findings are based on a small set of frequent and rare words. More research is needed involving a much larger set of words, including longer words.

- It may be that 10–12 weeks of classroom learning is not long enough to observe large changes in adult learners' speech perception ability. A full year (or longer) of classroom learning may reveal different results.
- The ability to perceive second language speech for learners of Mandarin enrolled in language classes is likely changing at small but measurable levels.

How to cite this summary: Wiener, S. (2019). Do classroom learners of Mandarin Chinese improve in their speech perception? *OASIS Summary* of Wiener et al. in *Language Learning*. https://oasis-database.org