Studies in Second Language Acquisition, **43** (2021), 155–180
doi:10.1017/S0272263120000418

Research Article

EFFECTS OF MULTITALKER INPUT AND INSTRUCTIONAL METHOD ON THE DIMENSION-BASED STATISTICAL LEARNING OF SYLLABLE-TONE COMBINATIONS

AN EYE-TRACKING STUDY

Seth Wiener *

Carnegie Mellon University

Kiwako Ito

University of Newcastle

Shari R. Speer

The Ohio State University

Abstract

To test the effects of talker variability and explicit instruction on the statistical learning of lexical tone, 80 monolingual English listeners were taught an artificial language that mimicked Mandarin's asymmetric distribution of syllable-tone co-occurrences. Training stimuli consisted of either speech from one talker or speech from four talkers. Participants were either never instructed or explicitly taught associations between phonemes (CVs), tones, and nonce symbols across four consecutive days. Learning was assessed by the accuracy of mouse clicks and eye movements to visual nonce symbols. Critical trials induced competition between the target symbol, which matched the acoustic input, and a competitor symbol that had a statistically more probable tone (but mismatched the acoustic input). Eye fixations indicated that participants were sensitive to syllable-tone co-occurrence probabilities even without explicit instruction of tone. The degree to

The experiment in this article earned an Open Materials badge for transparent practices. The materials are available at https://osf.io/pxd5f

This work was funded in part by a Doctoral Dissertation Research Improvement Grant from the National Science Foundation (BCS-1451677). We thank Marjorie Chan, Chao-Yang Lee, Mineharu Nakayama, Jessie Nixon, and the anonymous reviewers for their feedback on earlier versions of this work. We are especially grateful to James Brennan for his help with data collection.

*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. E-mail: sethw1@cmu.edu

which statistical knowledge was used to recognize words appeared to increase when participants processed more variable speech.

INTRODUCTION

The process of spoken word recognition requires the listener to match acoustic input to existing linguistic representations in memory. As a listener perceives a word's initial acoustic information, multiple lexical candidates are concurrently activated and compete against one another for selection (see Cutler, 2012 for a review). During this early stage of lexical processing, a listener may also draw on knowledge of statistical regularities about the sound patterns of words (Dahan et al., 2001; McQueen & Cutler, 2010). For a second language (L2) learner, spoken word recognition in the L2 requires learning novel speech cues and building the statistical knowledge of how often those multidimensional cues occur and co-occur with other cues to form sound categories, phonotactics, and words (e.g., Ellis, 2002, 2011; Escudero & Boersma, 2004; Hayes-Harb, 2007; Holt & Lotto, 2006; Pajak et al., 2016).

Previous L2 speech learning research investigating adult learners' acquisition of novel speech sounds has identified challenges stemming from a late age of acquisition, a perceptual system attuned to first language (L1) input, and limited L2 input in a classroom setting (see Colantoni et al., 2015 for a review). This research has established that L2 learners often struggle to phonetically discriminate novel speech sounds, particularly those produced by unfamiliar talkers whose acoustic realizations of speech may differ considerably from realizations previously encountered (e.g., Barcroft & Sommers, 2005, 2014; Hardison, 2003; Lecumberri et al., 2010; Lively et al., 1993; Wade et al., 2007). This inability to accurately discriminate new L2 speech sounds may cause learners to activate spurious lexical competitors, which ultimately results in slower or inaccurate L2 spoken word recognition (Weber & Cutler, 2004, 2006).

Given the challenges associated with recognizing L2 speech, statistical information about the probability of novel L2 speech cues may be highly beneficial for learners. In other words, information about the likelihood of sound occurrences and co-occurrences may facilitate word recognition in the absence of sophisticated L2 perception. Here we examine this claim with respect to a specific type of input learning, "dimension-based statistical learning" (Idemaru & Holt, 2011, 2014, 2020), which reflects listeners' ability to track multidimensional speech cues such as the statistical regularity of CV strings (where C and V refer to consonants and vowels, respectively) and the acoustic dimensions that define language units larger than the phoneme. Because talker variability can affect L2 learning (e.g., Barcroft & Sommers, 2005), we also examine how variability in the speech signal affects naïve listeners' reliance on statistical information for L2 spoken word recognition. We additionally clarify whether explicit instruction on particular speech cues (e.g., Saito, 2011, 2015) is beneficial to L2 learners. An investigation into the effect of explicit instruction on speech sounds may also benefit L2 educators, particularly those who teach speech sounds that are typically difficult for adult learners to acquire.

To address these questions, this eye-tracking study tests (1) whether adult listeners unfamiliar with lexical tone learn statistical regularities of syllable-tone co-occurrences,

(2) whether multitalker speech affects the reliance on statistical information for online L2 word recognition, and (3) whether explicit instruction on lexical tone contours facilitates this statistical learning.

BACKGROUND AND MOTIVATION

DIMENSION-BASED STATISTICAL LEARNING OF SYLLABLE-TONE CO-OCCURRENCES

Mandarin Chinese (hereafter "Mandarin") provides a relatively consistent mapping from the word to the written form to the morpheme to the spoken syllable (Myers, 2010; Packard, 1999, 2000). As an example, the word "horse" can be written with one character, 马, which corresponds to one morpheme, and which consists of one syllable: *ma*. Hereafter, we use "syllable" to refer to a segmental string irrespective of its tone. Because a syllable-tone combination like *ma3* is the minimal Mandarin morpheme or word, we use "word" hereafter to refer to a specific syllable-tone combination.

Each of the nearly 400 (C)V(C) Mandarin syllables varies in terms of its token frequency where the frequency of an item is inversely proportional to its frequency rank in a natural corpus (e.g., the tenth most likely word appears .1 times as often as the most likely one; DeFrancis, 1986; Duanmu, 2007, 2009; Zipf, 1935, 1949). In Mandarin, for example, the 30 most frequent syllables account for roughly 50% of the tokens in a 46.8-million-character speech corpus (SUBTLEX-CH: Cai & Brysbaert, 2010). In contrast, the 100 least frequent syllables account for approximately 1% of the corpus' tokens. Hereafter, references to frequency signify "token frequency." Mandarin word recognition thus involves repeatedly activating a small subset of highly frequent syllables and occasionally activating mid- to low-frequency syllables. Native speakers learn from this asymmetric distribution of syllables in speech and tend to perceive and produce frequent syllables faster than infrequent syllables (Chen et al., 2002, 2003; Zhou & Marslen-Wilson, 1994) as do classroom L2 learners (Wiener et al., 2019; Wiener & Lee, 2020).

In speech, each Mandarin syllable can be produced with up to four different lexical tones (Gandour, 1983; Howie, 1976). Figure 1 plots the four tones with their canonical fundamental frequency (F0) contours when spoken in isolation. Importantly, the segmental sequence composing each syllable constrains the probability of the four tones. For example, *gei* only forms a Mandarin word with the dipping third tone as *gei3*. *Gei1*, *gei2*, and *gei4* are all nonwords. Another example is the syllable *zhou*, which forms a word with all four tones. Yet, roughly 90% of all spoken *zhou* tokens in SUBTLEX-CH appear with the first tone as *zhou1*. L1 Mandarin speakers' linguistic knowledge therefore includes which syllable-tone combinations correspond to words, which combinations are nonwords, and which combinations are most likely to occur in speech (Fox & Unkefer, 1985; Wang, 1998; Wiener & Turnbull, 2016).

Such statistical information directly contributes to L1 Mandarin word recognition. In an eye-tracking study, Wiener and Ito (2015) demonstrated that native Mandarin speakers draw on this syllable-tone co-occurrence probability information during the early stages of spoken word recognition. Participants were simultaneously presented with an array of four frequency-controlled Chinese characters and an auditory stimulus corresponding to one of them. The experimental manipulation involved a target (e.g., *zhou2*) and a tonal

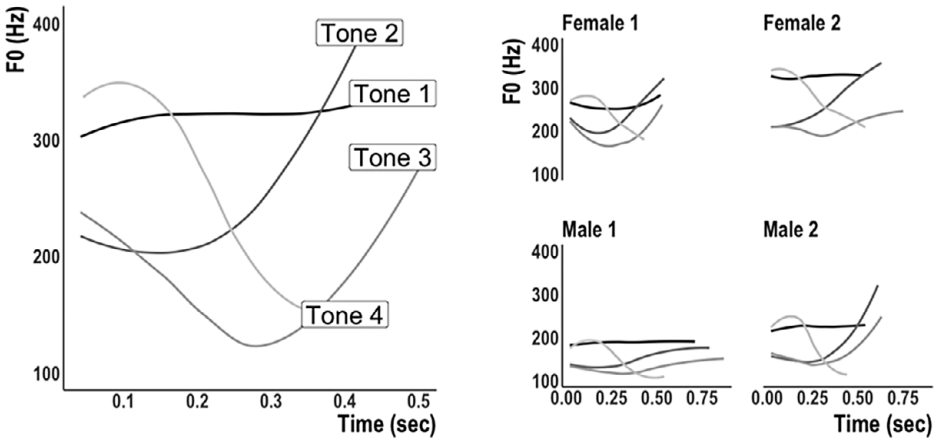

FIGURE 1. Four Mandarin tones spoken in isolation by a native female talker (left). Artificial language tones spoken in isolation by four talkers (right).

competitor, which shared the target's syllable but differed in tone (e.g., *zhou1*). Upon hearing a syllable's initial acoustic information, such as the onset of *zhou*, listeners initially looked to the character with the most probable tone (e.g., *zhou1*). For infrequent syllables, these predictive looks regularly occurred even if the tone of the competitor was incongruent with the incoming acoustic information (e.g., *zhou2*, which starts with a distinctively lower onset F0 than *zhou1*; see Figure 1). This effect of tonal probability, however, was short-lived—starting from roughly 300 ms to 700 ms after the onset of the syllable—and looks to the highly probable competitor were significantly higher only for infrequent target syllables. Wiener and Ito, consistent with previous research on homophony (e.g., Luce & Pisoni, 1998; Vitevitch & Luce, 1998, 1999; Vitevitch et al., 1999), claimed that tone is less informative for the identification of highly frequent Mandarin syllables with a large number of homophonous morphemes. As an example, *yi* —the ninth most frequent syllable in SUBTLEX-CH—with the falling fourth tone can be written with nearly 100 semantically and orthographically distinct characters (Yin, 1984). Tonal information may not help select specific *yi* lexical candidates because word identification for frequent syllables like *yi* typically requires additional context or morphemes (Packard, 1999, 2000; Zhou & Marslen-Wilson, 1994, 1995). In contrast, tone is more informative for the identification of infrequent syllables like *zhou*, which appear with far fewer homophones in sparse neighborhoods (e.g., Chen et al., 2009; Li & Yip, 1998; Wiener & Ito, 2016).

Thus, L1 Mandarin speakers make use of syllable-tone dimension-based statistical information for online word recognition. Recent findings indicate that intermediate L2 classroom learners familiar with lexical tone appear to exhibit similar dimension-based statistical learning of a CV-syllable and tone (Wiener et al., 2018). L1 English adults enrolled in an intermediate L2 Mandarin language course took part in a multiday artificial tonal language-learning task (e.g., Caldwell-Harris et al., 2015; Wewalaarachchi et al., 2017). The syllable-tone input was designed to mimic Mandarin's varying syllable frequency and conditional tone probabilities. After four consecutive days of training,

L2 learners demonstrated comparable statistical experience-based early looks to more probable targets given the perceived syllable. Like the native speakers tested in Wiener and Ito (2015), the L2 learners' predictive looks were primarily observed for low-frequency syllables. These looks were rapidly shifted to the correct symbols if the initial probability-based prediction was incongruent with the acoustic information.

What remains unclear is whether knowledge of lexical tone is necessary for learners to track its statistical regularities. Statistical learning of a nonnative speech cue may only emerge only once learners are consciously aware of the novel cue's importance in word recognition. The present study serves as a follow-up to Wiener et al. (2018) to examine whether adult listeners unfamiliar with lexical tone, that is, the typical adult learner in a beginner Mandarin classroom language course, can learn statistical regularities involving syllable-tone co-occurrences with conscious attention.

TALKER VARIABILITY AND L2 TONAL WORD RECOGNITION

The present study additionally investigates how variability in the speech signal affects L2 learning and word recognition. With respect to Mandarin tones, previous research established that learners struggle to distinguish between low-F0-onset tones 2 and 3 and high-F0-onset tones 1 and 4, even after extended L2 classroom experience (e.g., Hao, 2012, 2018; Leather, 1983; Pelzl, 2019; Pelzl et al., 2019). For many L2 learners, this perceptual confusion increases when faced with multitalker speech. Figure 1 (right) illustrates how the same four syllable-tone productions vary across talkers in their F0 onset, contour, and range, a circumstance that could lead to tonal errors among L2 learners (e.g., Chang & Bowles, 2015; Lee et al., 2009, 2010, 2013; Qin et al., 2019; Shen & Lin, 1991).

Wang et al. (1999) first explored whether exposure to more varied, multitalker input improves beginner L2 Mandarin learners' tone categorization. Participants heard Mandarin syllable-tone words from four talkers and were asked to categorize the speech according to its tone type. After 2 weeks of multitalker training, participants improved in their overall tone categorization. This learning even generalized to new tokens and speakers. Yet, follow-up multitalker studies (Perrachione et al., 2011; Sadakata & McQueen, 2014) that controlled for individual tonal aptitude found that only learners with strong perceptual abilities benefited from greater talker variability—those learners with weak abilities saw little improvement (though see Dong et al., 2019 for conflicting results).

Whereas greater talker variability may improve L2 tonal categorization, limited evidence suggests it may also cause uncertainty during L2 spoken word recognition. The L2 learners tested on Wiener et al. (2018)'s artificial language demonstrated a greater reliance on tonal probability information when talker variability increased. Learners trained and tested on input from four talkers across the four days of training looked more to probable targets and recovered more slowly from incorrect predictions than those trained and tested on input from the same talker. The authors argued that as learners' uncertainty about the tone category increased with multitalker input, they relied more on their statistical knowledge of syllable-tone co-occurrence to predict the intended target.

Thus, for intermediate L2 learners still learning to categorize tones, reliance on tonal probability information increased when faced with greater acoustic variability in the

input. Because neither Wang et al. (1999) nor Wiener et al. (2018) tested truly naïve listeners, the current study aims to clarify how multitalker input affects naïve listeners' learning of syllable-tone words and their reliance on statistical information for spoken word recognition.

STATISTICAL LEARNING OF TONE AND EXPLICIT INSTRUCTION

The present study looks for evidence to support an innate statistical learning mechanism that facilitates language learning and processing using a process of extracting distributional information for sounds occurring in the continuous speech signal (Gómez & Gerken, 2000; Maye et al., 2002; Saffran, 2003). Whereas the majority of research on the statistical learning of L2 speech sounds has focused on consonants and vowels (e.g., Escudero et al., 2011; Wanrooij et al., 2013), we test whether this mechanism extends to the statistical learning of lexical tones.

Previous research on statistical learning of tone suggests that listeners can learn distributional regularities of words with tones, but it remains unclear how tonal information is represented in leaners' lexicons. A study by Wang and Saffran (2014) paired nine unique artificial tones with nine CV syllables to form three trisyllabic tonal words (e.g., *bidatu* with each syllable carrying a different tone contour). Adult participants were instructed to listen to a 9-minute continuous stream of speech containing these words. Learners could therefore track regularities involving the syllable-only, the tone-only, or the syllable-tone combinations as "words." After the familiarization phase, participants heard a word from the training along with a new word that reversed the order of the syllables (e.g., *tudabi*). Participants had to indicate which of the two CV strings sounded more familiar. Monolingual English listeners performed the task at chance, whereas monolingual Mandarin listeners successfully discriminated previously heard words from nonwords (though the two monolingual groups did not differ statistically). Interestingly, two groups of bilinguals (Mandarin–English bilinguals and English and a nontonal L2 bilinguals) both discriminated words from nonwords better than the two monolingual groups. The authors concluded that experience with multiple languages—irrespective of whether the languages are tonal—resulted in an increased ability to discover patterns in new stimuli.

A follow-up study by Potter et al. (2017) used the stimuli and task of Wang and Saffran (2014) to test two groups of listeners: a monolingual English control group and an L1 English group currently learning Mandarin. The two groups discriminated words from nonwords with nearly identical accuracy levels of 55%. When tested 6 months later, the participants enrolled in the L2 Mandarin course improved to 66% while the monolingual control group remained slightly above chance (53%). Potter et al. argued that this improvement demonstrates that the L2 learners transferred their classroom experience to the statistical learning task. However, it is unclear what cues the participants were tracking: syllable regularities, tone regularities, or syllable-tone regularities.

Critically, the stimuli used in Wang and Saffran (2014) and Potter et al. (2017) presented syllables with the same tone, that is, there were no phonotactically identical sequences distinguished only by tone. Therefore, tones from their artificial language did not signal lexical contrast. Evidence from neuroimaging and eye-tracking data (Malins & Joanisse, 2010, 2012) suggests that lexical tone processing, where sound is associated to

symbol and thus used for word identification, may fundamentally differ from nonlexical tone processing. The present study manipulates the tonal distribution separately from the syllable distribution to clarify whether novice listeners can extract statistical regularities of tone-syllable co-occurrences and use such information for online word recognition involving sound-symbol pairs. As we outline in the following text, this approach involves training participants on a larger set of items to set up the distribution of tone-syllable correspondences, and then testing participants with a smaller critical set of items to see the effect of frequency of co-occurrence.

The present study additionally aims to test the effect of explicit instruction on lexical tone acquisition and statistical learning. Previous research reports that adult performance in statistical learning tasks improves when explicit instructions to attend to the stimuli are given (e.g., Ong et al., 2015; Saffran et al., 1997; Turk-Browne et al., 2010). Ong et al. (2015) trained naïve L1 English listeners on Thai tonal minimal pairs and measured participants' ABX discrimination before and after training. Participants improved their statistical learning of tone, but only if their auditory attention was encouraged using a cover task. Thus, statistical learning may be more effective if learners are instructed to attend to specific acoustic dimensions that matter for learning novel words. Ong et al.'s finding may also explain why the intermediate L2 participants tested in Wiener et al. (2018) were able to track syllable-tone co-occurrences: These learners had more than a year of explicit Mandarin classroom training and were clearly aware of tonal cues in the stimuli. However, as with the Potter et al. (2017) and Wang and Saffran (2014) stimuli, the Ong et al. stimuli were not lexical in nature and participants' learning was assessed with phonetic discrimination tasks only. Thus, listeners may not have categorized tones as lexical cues. The present study makes the lexical function of tone overt to listeners using sound-symbol pairs, and tests whether explicit instruction facilitates acquisition of lexical tone and its statistical regularities.

RESEARCH QUESTIONS

RQ1. To what extent does listeners' experience with lexical tone affect their dimension-based statistical learning of syllable-tone co-occurrences? Can adult listeners unfamiliar with lexical tone learn syllable-tone statistical regularities?

RQ2. To what extent does multitalker input modulate the reliance on statistical information for online L2 spoken word recognition? Do learners exposed to multiple talkers rely on statistical knowledge to a greater degree than learners exposed to only one talker?

RQ3. To what extent does explicit instruction of lexical tone facilitate dimension-based statistical learning? Can naïve listeners who are not explicitly aware of lexical tone still track syllable-tone statistical regularities?

EXPERIMENT

METHOD

Participants

Eighty native speakers of American English with no previous experience in Mandarin or another tonal language participated in the study (M_{age} = 20.9, SD = 4.0). All participants

were undergraduates at a Midwestern university and had normal or corrected vision, normal speech and hearing, and no formal musical training. Tonometric (Mandell, 2015) was used to control for pure pitch perception.[1] All participants had previously studied at least one nontonal European language during secondary school (e.g., French, German, Italian, Russian, or Spanish). Participants were asked to self-rate their L2 abilities on a 4-point Likert scale (1: beginner; 4: fluent: $M_{speaking} = 2.1$, $SD = 0.8$; $M_{listening} = 2.1$, $SD = 0.7$). No participant self-rated as a fluent speaker or listener of an L2 or studied another language at the time of testing. All participants were paid for their time.

Materials

Eight consonants /p, p^h, f, k, k^h, m, ts, ɹ/ and six vowels /a, i, ə, ia, iu, ai/ were used to construct 24 Mandarin-like CV syllables (see Supplementary Materials for full stimuli). Each syllable was paired with four tonal contours similar to those in Mandarin (Figure 1). Only 82 of the possible 96 syllable-tone co-occurrences were used to approximate the natural rate of syllable-tone gaps within the Mandarin lexicon (see Myers, 2002, 2010; Wang, 1998). To introduce homophony within the artificial language, 48 syllable-tone homophones were added, resulting in 130 total nonce words—a rate of homophony approximating that of spoken Mandarin (see Duanmu, 2007, 2009). Each nonce word was given a unique symbol (Figure 2). These nonce symbols—like Chinese characters—served to disambiguate homophonous syllable-tone co-occurrences (see Wiener, 2015 for details on symbol creation). This meant that some syllable-tone combinations mapped to multiple symbols whereas other syllable-tone combinations only mapped to one symbol.

To manipulate the statistical distribution of syllables and tones within the artificial language, 66 of the 130 items in the language served as fillers. The remaining 64 items served as critical items. Among the 64 critical items, two factors—syllable token frequency and syllable-conditioned tonal probability—were crossed to make four within-subject stimulus conditions. Syllable frequency was manipulated by increasing or decreasing the number of syllable tokens to which participants were exposed, irrespective of tone. Thirty-two of the 64 test items had high syllable frequency (F+: 28 tokens per day for a total of 896 tokens) while the other 32 items had low syllable frequency (F−: 16 tokens per day for a total of 512 tokens). For each syllable, a particular tone type was

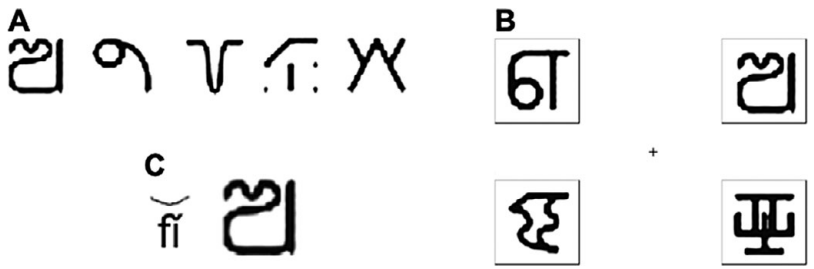

FIGURE 2. A: Example of five *fi3* homophones. B: Sample 4-AFC slide; target (*fi1*, top-left), tonal competitor (*fi3*, top-right), rhyme competitor (*ri1*, bottom-right), and distractor (*ka2*, bottom-left). C: Example of explicit training slide for *fi3*.

assigned to be the most probable (P+; occurring in 50% of the tokens) while another tone type served as the least probable (P−; occurring in roughly 10% of the tokens); the other two tones had identical mid-range probabilities. These frequency and probability values approximated those used in Wiener and Ito (2015) while providing a reasonable amount of input for participants to learn the artificial language based on pilot results. Figure 3 illustrates the input to participants. Each syllable-tone input varied in the number of symbols to which it mapped, which presumably affected participants' certainty about the identity of the target.

These frequency and probability manipulations resulted in four within-subject stimulus conditions: F+P+, F+P−, F−P+, F−P−. Filler homophones were used to manipulate tonal probabilities, which allowed for the target test items to visually appear the same number of times. For example, on each day of training and testing, the syllable-tone combination *fi3* was presented 10 times using the four rightmost nonce symbols in

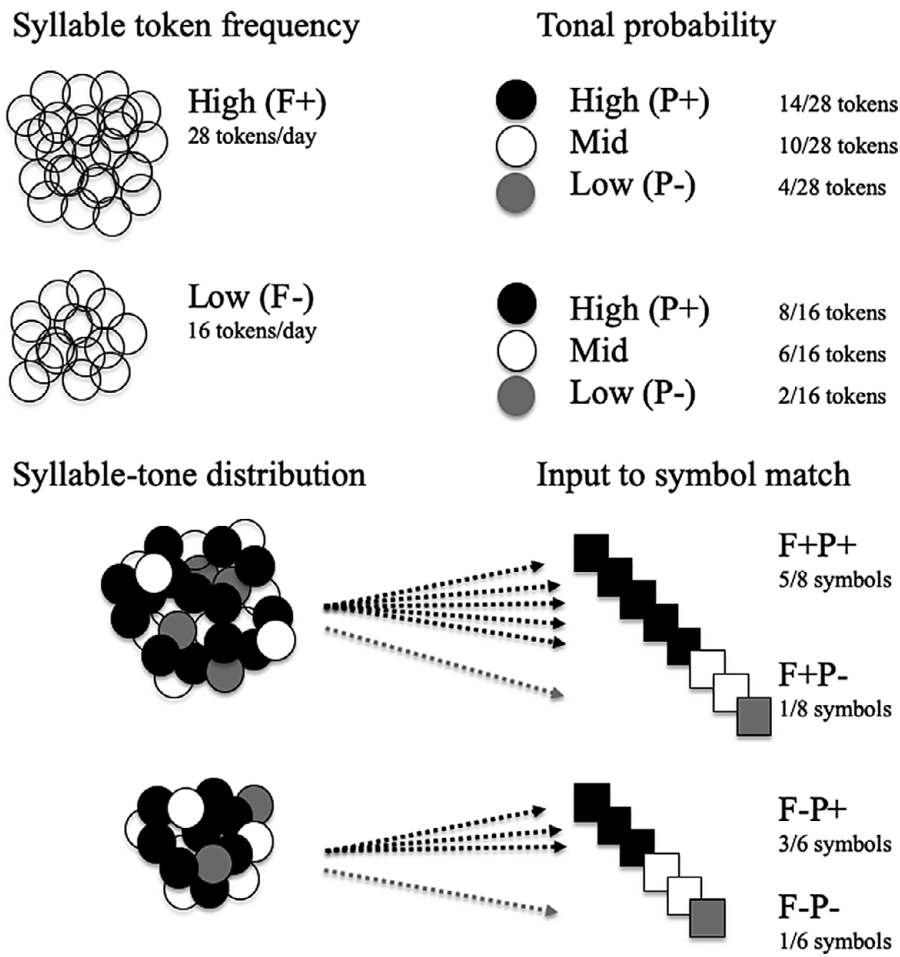

FIGURE 3. Illustration of syllable-tone input and match to symbols.

Figure 2A. By continually showing participants these four symbols throughout the experiment, the probability of the syllable *fi* appearing with tone 3 greatly increased. The leftmost nonce symbol in Figure 2A functioned as the test item and visually appeared four times each day, which is the same number of times as the three respective critical item symbols for *fi1*, *fi2*, and *fi4*. Thus, the critical test items' symbols appeared with identical frequency while the phonological occurrence of a syllable with a particular tone varied through manipulation of filler homophone symbols.

All auditory stimuli were recorded by four (two male and two female) native Mandarin talkers at 16 bits/44,100 Hz. Figure 1 (right) shows an example of *fi1*, *fi2*, *fi3*, and *fi4* produced by each talker. Acoustic analysis of the stimuli confirmed previously reported temporal differences; tones 2 and 3 were longer in duration than tones 1 and 4 (Ho, 1976; Moore & Jongman, 1997; Zee, 1980). Mean duration of each tone type was comparable across the four talkers: tone 1 ($F(3,96) = 0.166$, $p = 0.69$); tone 2 ($F(3,96) = 0.825$, $p = 0.37$); tone 3 ($F(3,96) = 3.28$, $p = 0.08$); and tone 4 ($F(3,96) = 0.114$, $p = 0.73$).

Design and Procedure

To control for the potential influence of particular sound-symbol pairs on learning, participants were randomly assigned to one of two presentation lists that counterbalanced sound-symbol mapping. Participants took part in 30-minute training-testing sessions on four consecutive days thereby allowing for overnight consolidation of word learning (Qin & Zhang, 2019). Participants performed the tasks in the same order each day: passive listening, shadowing, naming (with feedback), and 4-alternative forced-choice (4-AFC) with eye-tracking (with feedback).[2] Figure 4 shows the daily training and testing routine.

Participants first completed a self-paced passive listening task (131 trials). They were seated in a sound-attenuated booth with a computer monitor and given headphones. A symbol and its audio label were simultaneously presented. Participants were told to remember the sound-symbol pair and mouse click to advance to the next trial. Participants next shadowed the speech (131 trials; because the results are beyond the scope of this article, see Wiener et al., 2020 for discussion). After a sound-symbol pair was presented, participants were asked to repeat the perceived word as clearly and accurately as possible. The symbol remained on screen until the participant mouse clicked to advance to the next trial. After completing the shadowing task, participants performed a naming task (64 trials). These trials, which only tested the 64 critical items in the language, first presented a symbol on screen. Participants were told to produce the symbol's syllable-tone as accurately as possible and to guess if they were unsure. The symbol remained on screen until the participant mouse clicked, at which point feedback about the correct audio label was presented using headphones. Participants ended each day of training with the 4-AFC task, which is the focus of this article.

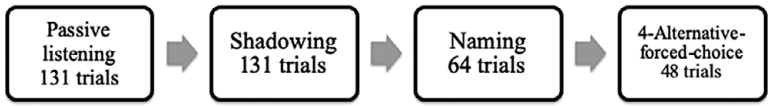

FIGURE 4. Daily training and testing routine.

In the 4-AFC task, participants' eye movements were first calibrated using the Clearview 5-point calibration program. At the start of each trial, participants fixated on the center of the screen while listening to a carrier phrase "I will say…" ("*wo3 yao4 shuo1*" spoken in Mandarin), followed by simultaneous presentation of four symbols and the target symbol's audio label, that is, a slide was not previewed. Participants were told to click on the symbol that matched the perceived audio while their eye movements were continuously recorded at 50 Hz using a Tobii 1,750 system. After clicking, correct targets were highlighted with a red box so that feedback could guide learning. There were 48 total trials (16 target trials with four in each stimulus condition and 32 filler trials) with a 2 second intertrial interval.

The 16 target trials (consisting of symbols only from the 64 critical items) showed the target and three other trained test items: a tonal competitor, a rhyme competitor, and a distractor. Figure 2B shows a sample trial from the high token frequency, low-probability condition (F+P−) with *fi1* (P−) as the target (top left). The three other on-screen items included a tonal competitor, which shared the same syllable but had the opposite tonal probability (e.g., P+ *fi3*; top right); a rhyme competitor, which shared the same vowel and tone but had a different onset (e.g., *ri1*; bottom right); and a distractor, which had a wholly dissimilar syllable and tone (e.g., *ka2*; bottom left). Position of the target and competitors was fully counterbalanced across days and trials.

To test the effect of talker variability, half of the participants were trained and tested on a single female talker (single-talker condition). Participants in this condition heard speech from Female 1 during all four tasks: the 4-AFC test consisted of a familiar syllable-tone exemplar spoken by a familiar talker. The other half of the participants were trained and tested on four different talkers (multitalker condition): the 4-AFC test consisted of a familiar syllable-tone learned in the first three tasks but spoken by a particular talker for the first time. For instance, participants heard *fi1* in the first three training tasks spoken by Male 1. Participants were then tested in the 4-AFC task on Female 2's production of *fi1*.

To test the effect of instruction, half the participants were given no instructions regarding lexical tone whereas the other half were explicitly trained. At the start of each daily training session, participants in the explicit training condition took part in a 5-minute, self-paced computerized lesson on the four tones using the *ma* tone quadruplets as practice sounds. This lesson was presented in English and emphasized the four tones' F0 contours following modern L2 classroom pedagogy for introducing isolated monosyllables (e.g., Shen, 1989; Xing, 2006; Yang, 2017). Participants were first told to "pay attention to the pitch or tone of the syllable." Participants then simultaneously heard a syllable tone while its *pinyin* was displayed onscreen with tone diacritics. After participants mouse clicked on the screen, the syllable-tone was presented again over headphones while the syllable's rendered F0 contour was displayed on screen (generated through Praat; Boersma & Weenink, 2016). Thus, participants were trained to associate the four F0 contours with the four tonal categories. At the conclusion of the lesson, participants were told that the "pitch or tone of the syllable is important for symbol learning."

For the explicit group, nonce symbols were displayed alongside the symbol's syllable-tone label in *pinyin* and the tone's rendered F0 contours during the passive learning and shadowing phases of daily training. Figure 2C shows an example for the *fi3* target

containing the *pinyin* and tone contour. Participants in the no-instruction group were only shown the nonce symbol. Hereafter, we refer to the between-subject variables talker variability (single-talker/multitalker) and phonetic instruction (explicit/nonexplicit) as *training conditions* while the within-subject variables syllable token frequency (F+/F−) and tonal probability (P+/P−) as *stimulus conditions*.

Predictions

We predict higher 4-AFC mouse-click accuracy for those participants trained with explicit instruction to attend to F0 contours, which should increase awareness of the phonological role of tones (e.g., Chun et al., 2015; Godfroid et al., 2017; Liu et al., 2011; Saito & Wu, 2014). Explicit training may interact with talker variability such that explicit tone instruction is effective only for highly familiar or low variability speech. In this outcome, participants explicitly trained on single-talker input may demonstrate the highest overall mouse-click accuracy. Participants trained on multitalker input may show little to no effect of explicit instruction given the challenge associated with multitalker tones (e.g., Lee et al., 2010; Wang et al., 1999).

Participants' eye movements within the first 1,000 ms will reveal whether syllable-tone co-occurrence probabilities were used during online spoken word recognition. In particular, we test whether participants activate the most probable (P+) tone upon hearing a syllable's initial acoustic information (RQ1). This will be shown in anticipatory fixations to the more probable (P+) symbol as compared to the less probable (P−) symbol. If talker variability challenges participants' perception of tone (e.g., Wiener et al., 2018), participants exposed to multitalker input may rely on probabilistic information to a greater degree and look to the symbol with the more probable (P+) tone (RQ2). If explicit instruction facilitates statistical learning of tone distributions, participants explicitly trained on tone may respond with more anticipatory eye movements to more probable (P+) visual candidates (RQ3).

RESULTS

4-AFC MOUSE-CLICKS

Mean 4-AFC mouse-click accuracy was calculated for each of the four groups for each day. Figure 5A shows the changes in accuracy (with 95% confidence intervals) across the 4 days of testing (x-axis). Participants performed above chance (.25) on Day 1 regardless of the training conditions (bar color) and demonstrated overall average daily improvements with each training/testing session. Figure 5B plots Day 4 results by stimulus (x-axis) and training conditions (bar color). Participants in the two single-talker training conditions had roughly equal mean accuracies across the four stimulus conditions. In contrast, participants in the two multitalker training conditions had the lowest mean accuracy for low-frequency low-probability (F−P−) targets and the highest mean accuracy for low-frequency high-probability (F−P+) targets.

To test the effects of training conditions and stimulus conditions on the mouse-click accuracy, Day 4 (i.e., the final day of testing) mouse-click results were analyzed using mixed-effect logistic regression models with the *lme4* package (Bates et al., 2015) in R (version 3.6.1; R Core Team, 2019). The inclusion of fixed and random effects of all

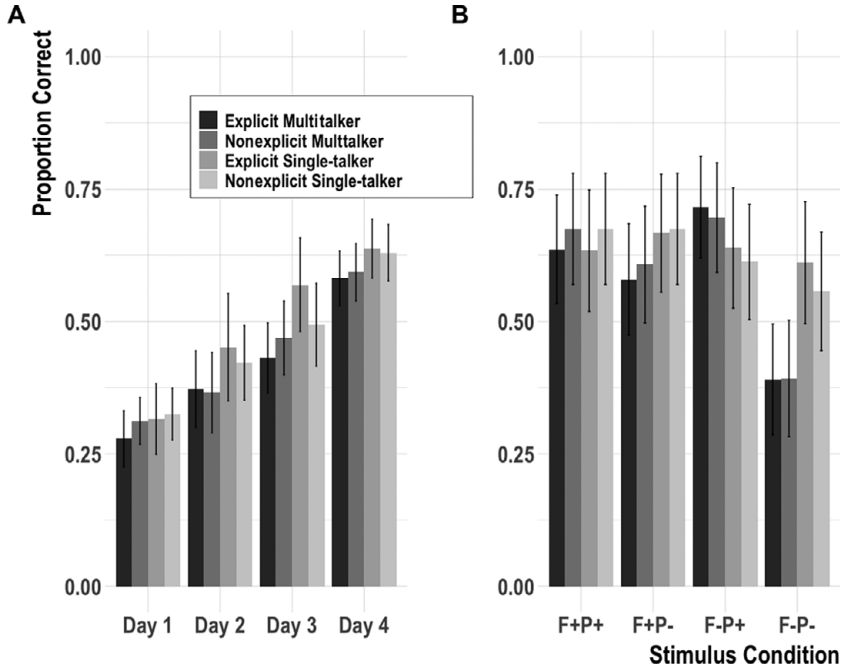

FIGURE 5. A: Mean daily 4-AFC accuracy by training conditions. B: Mean Day 4 4-AFC accuracy by stimulus and training conditions. Error bars in both figures indicate 95% confidence intervals.

TABLE 1. Summary of mixed-effect regression on Day 4 4-AFC mouse-click accuracy

	Estimate	*SE*	*z*	*p*
(Intercept)	0.46	0.06	7.50	< .001
Syllable frequency	0.14	0.05	2.35	0.02
Tonal probability	0.21	0.05	3.62	< 0.001
Freq:Prob	−0.17	0.06	−2.85	0.004
Talker variability:Prob	−0.19	0.06	−3.22	0.001

Note: High frequency, high probability, and single-talker were all coded as 1.R code: glmer(accuracy ~ syllable frequency × probability + talker variability: probability + (1+variabilitylitem) + (1+syllable frequency × probability|subject), family = "binomial").

models in this article were evaluated by first building the maximally appropriate model with random intercepts and slopes: glmer(4-AFC accuracy ~ talker variability × syllable frequency × probability × instruction + (1 + instruction × talker variabilitylitem) + (1 + syllable frequency × probability|subject), family = "binomial"). Next, effects that did not improve the model fit as assessed through the *lmertest* package (Kuznetsova et al., 2017) were removed. See Supplementary Materials for power analyses carried out using the *simr* package in R (Green & MacLeod, 2016). All models in this article treated fixed effects as contrast coded variables (1, −1). The final 4-AFC logistic regression model did not include instructional manner as a main effect (χ^2 = 0.01, p = .92) or as part of any interaction (χ^2s < 2, ps > .1). Talker variability was not included as a main effect (χ^2 = 2.42,

$p = .11$), as a two-way interaction with syllable frequency ($\chi^2 = 0.02$, $p = .88$), or as a three-way interaction with syllable frequency and tonal probability ($\chi^2 = 3.11$, $p = .07$). See Table 1 for final model specification and output.

The model revealed (positive) main effects of syllable frequency and tonal probability and two separate (negative) two-way interactions between syllable frequency and probability, and between talker variability and probability. The difference in direction of these effects indicates that the main effects were not additive. For low-frequency (F−) items, there was an effect of probability and for low-probability (P−) items, there was an effect of frequency. In other words, no difference was found between probability conditions for F+ items ($p = .24$)[3]; for F− items, high-probability tones (P+) were identified more accurately than P− tones ($p < .01$). For P+ tones, no difference was found between frequency conditions ($p = .12$); for P− tones, F+ items were identified more accurately than F− items ($p < .01$). For single-talker speech, no difference was found between probability conditions ($p = .74$); for multitalker speech, P+ tones were identified more accurately than P− tones ($p < .001$).

To summarize, tonal probability was useful for the identification of low-frequency syllables only (F−P+, F−P−). No difference due to tonal probability was found for items containing high-frequency syllables (F+P+, F+P−). Moreover, this effect of tonal probability was primarily driven by participants exposed to multitalker input. Participants exposed to single-talker input did not identify P+ items more accurately than low-probability targets. These results suggest that participants unfamiliar with tone languages can learn syllable-tone statistical regularities (RQ1), but they rely on this information to a greater extent when the stimuli contain high talker variability (RQ2). Explicit instruction had no apparent effect on participants' ability to extract statistical information (RQ3).

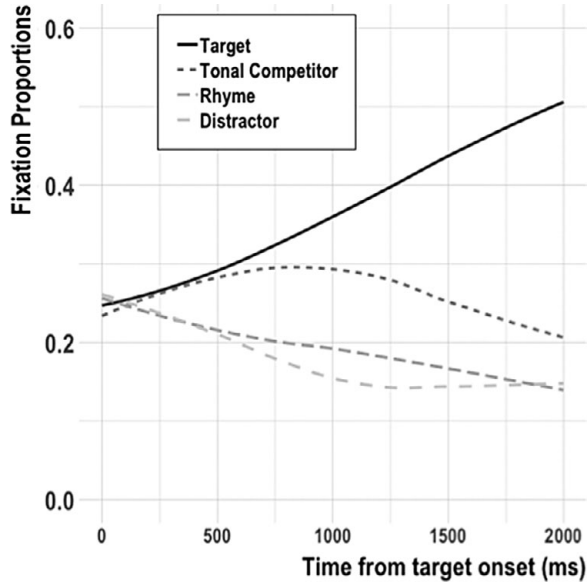

FIGURE 6. *Loess*-smoothed grand mean fixation proportions from the onset of the target word across 4 days of testing.

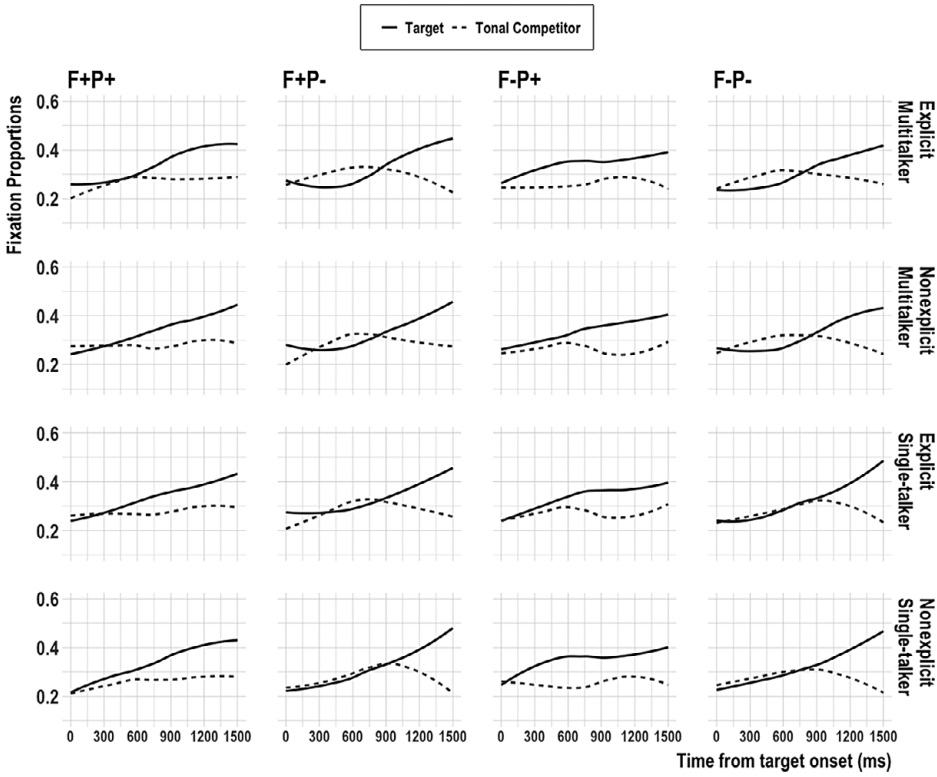

FIGURE 7. *Loess*-smoothed mean fixation proportions to Day 4 target and tonal competitor by stimulus and training conditions (0–1,500 ms).

Eye Fixations

To examine the overall time course of shifts in eye fixations and to determine the sizes of analysis windows, we first plotted the *loess*-smoothed (Cleveland, 1979) grand mean fixations to the target and competitors across all training and stimulus conditions on all 4 days (Figure 6). Looks to each visual candidate began at chance (.25) at the critical word onset and looks to the target and tonal competitor diverged from looks to the rhyme competitor and distractor at about 250 ms after that. In accordance with previous studies that report that approximately 200 ms (from critical auditory input) is required to plan and execute an eye movement (e.g., Matin et al., 1993), the initial divergence reflects accurate detection of syllable onset. Looks to the tonal competitor peaked between 750 ms and 1,000 ms and dropped below chance (.25) by approximately 1,500 ms.

Figure 7 magnifies the 300–1,500 ms time window in which tonal competition was above chance. The figure plots the *loess*-smoothed fixation functions for the target and tonal competitor by training and stimulus conditions on Day 4. Both high-probability conditions (1st and 3rd columns) showed an early divergence between the target and tonal competitor, and the looks to the tonal competitor remained at or below chance after 300 ms, regardless of training condition. In these conditions where the target had

TABLE 2. Summary of weighted mixed-effect regressions on Day 4 target fixations

	Estimate	*SE*	*t*	*p*
300–500 ms:				
(Intercept)	−0.80	0.06	−12.28	< .001
Tonal probability	0.12	0.07	1.87	0.10
500–700 ms:				
(Intercept)	−0.62	0.05	−12.66	< .001
Tonal probability	0.12	0.05	2.48	0.03
700–900 ms:				
(Intercept)	−0.45	0.04	−12.33	< .001
Tonal probability	0.15	0.04	4.06	< .01
900–1,100 ms:				
(Intercept)	−0.35	0.03	−10.52	< .001
Tonal probability	0.08	0.03	2.26	0.05

R code: lmer(elog target ~ probability + (1|item) + (1 + probability|subject), weights = weight)

the most probable tone, looks to the less probable tonal competitor did not exceed looks to the target at any point in time. In contrast, the two low-probability conditions (2nd and 4th columns) showed looks to the tonal competitor at or above chance from 300 ms until 900 ms. For the two multitalker conditions (1st and 2nd rows), looks to the tonal competitor exceeded looks to the target competitor from roughly 300 ms until 900 ms.

To test whether looks to the target differed across training and stimulus conditions, the empirical logit (elogit) and the associated weights were calculated for 200 ms time intervals for each trial and participant. Four consecutive weighted mixed-effects models were tested from 300 ms until looks to the tonal competitor peaked at 1,100 ms. Model effect structure for each window followed the method outlined in the previous section by starting with the full model: lmer(elog target ~ talker variability × syllable frequency × probability × instruction + (1 + instruction × talker variability|item) + (1 + syllable frequency × probability|subject), weights = weight). The final model only contained tonal probability as a fixed effect. Syllable frequency, instructional method, and talker variability (and all interactions containing these variables) did not significantly improve the model's fit at any time window (χ^2s < 3, *p*s > .05). Given the number of models built, Bonferroni correction was used for all reported p-values. Table 2 reports the final model along with its output for each 200 ms window.

From 300 to 500 ms, a marginal effect of tonal probability was found. Participants looked to targets with high-probability (P+) tones at a marginally greater proportion than to targets with low-probability (P−) tones. From 500 to 700 ms and 700 to 900 ms, a main effect of tonal probability was found. Participants looked to P+ targets at a significantly higher proportion than to P− targets. From 900 to 1,100 ms, P+ targets were looked to at a marginally greater proportion than P− targets.

To confirm whether this reduction in looks to the P− target was due to participants looking to the more probable P+ tonal competitor, the weighted elogit of the tonal competitor was analyzed following the same procedure as outlined for looks to the target (Table 3). Once again, neither syllable frequency nor instructional method nor talker variability (nor any interactions) significantly improved the model's fit (χ^2s < 3, *p*s > .05).

TABLE 3. Summary of weighted mixed-effect regressions on Day 4 tonal competitor fixations

	Estimate	*SE*	*t*	*p*
500–700 ms:				
(Intercept)	−0.07	0.05	−12.42	< .001
Tonal probability	−0.12	0.06	−2.06	0.10
700–900 ms:				
(Intercept)	−0.64	0.05	−12.80	< .001
Tonal probability	−0.13	0.05	−2.62	0.04

R code: lmer(elog competitor ~ probability + (1|item) + (1 + probability|subject), weights = weight)

No significant effects of predictor factors were found in the 300–500 ms model. From 500 to 700 ms, participants looked to P+ tonal competitors at a marginally higher proportion than they looked to P− tonal competitors. From 700 to 900 ms, a main effect of tonal probability was found: participants looked to P+ tonal competitors at a higher proportion than to P− tonal competitors irrespective of syllable token frequency. No significant effects were found in the 900–1,100 ms window.

To summarize the eye-fixation results, from 300 to 1,100 ms, participants looked to high-probability (P+) candidates at a higher proportion than to corresponding low-probability (P−) candidates. These looks were independent of syllable frequency, talker variability, and instructional method. From 500 to 900 ms, participants looked at P+ targets and P+ tonal competitors at significantly higher rates than P− targets and P− tonal competitors across all training conditions (RQ1). Talker variability in the input (RQ2) and explicit instruction (RQ3) did not appear to affect the use of syllable-tone statistical information during online spoken word recognition.

DISCUSSION

This study examined whether adult listeners unfamiliar with lexical tone are able to track statistical regularities in syllable-tone co-occurrences in a manner similar to L1 and intermediate L2 Mandarin listeners. We additionally explored whether multitalker input increases the reliance on this statistical information during spoken word recognition and whether explicit instruction to attend to specific tone contours facilitates this statistical learning. We trained participants for four consecutive days on our artificial tonal language and took Day 4 eye movements as our measure of online predictive processing.

Results from the Day 4 4-AFC task indicated that tonal probability information was useful for identifying items containing low-frequency (F−) syllables only. No differences were found for items containing high-frequency (F+) syllables. Moreover, this effect of tonal probability was primarily driven by participants exposed to multitalker input. Participants in the single-talker condition did not show a statistically significant advantage in their mouse-click results for high tonal probability (P+) over low-probability (P−) items. Analyses of the Day 4 eye movements revealed that participants across all training conditions looked to P+ targets and tonal competitors at a significantly greater proportion than corresponding P− targets and tonal competitors from 500 to 900 ms. These eye movements were independent of instructional method or talker variability.

Taken together, our mouse-click and eye-movement results indicate that learners unfamiliar with lexical tone were able to track the relative frequency of CV syllables and the probabilities of such syllables co-occurring with particular F0 contours. This dimension-based statistical learning occurred independently of the variability in the speech input or the instructions used to teach tone. Early-stage L2 acquisition thus appears to involve not only tracking nonnative speech cues' distributions, such as phonetic continua and phonemic contrasts (e.g., Escudero et al., 2011; Hayes-Harb, 2007; Hayes-Harb & Masuda, 2008), but also multiple co-occurring cues that vary along acoustic-phonetic dimensions. Importantly, our results strengthen the claim that adult nonnative listeners can track tonal regularities at not just the phonetic level (e.g., Ong et al., 2017; Potter et al., 2017; Wang & Saffran, 2014) but also at the *lexical* level as in Mandarin.

Remarkably, this dimension-based statistical learning took place even without any explicit instruction of tone's acoustic characteristics. Unlike Ong et al. (2015), we found statistical learning of syllable-tone co-occurrences independent of instructional method. Participants trained explicitly on visualized tone contours and a *pinyin*-like romanization did not differ in mouse-click accuracy or proportion of eye fixations as compared to those participants who were not explicitly trained. Previous instructional method studies on nonnative tone acquisition typically used small, evenly distributed sets of syllable-tones with *pinyin*-like text (e.g., Showalter & Hayes-Harb, 2013), auditory-only stimuli devoid of visual cues (e.g., Godfroid et al., 2017), or tone-only discrimination tasks (e.g., Ong et al., 2015). In our study, the artificial language contained a large number of items (including syllable-tone homophones) and nonce symbols more closely resembling Mandarin. Our design thus may have better simulated the challenges involved in L2 acquisition (see Ettlinger et al., 2016 for discussion of artificial languages' external validity) by capturing the increased perceptual competition among new L2 sound categories and lexical candidates typically observed during spoken L2 word recognition (e.g., Escudero et al., 2008; Weber & Cutler, 2004, 2006). We note that the present study's Day 4 mean accuracy across training and stimulus conditions was approximately 60% with learners primarily making tonal mistakes. This is much lower than the posttraining ceiling performances typically observed with relatively small, symmetric stimuli syllable-tone sets (e.g., Wong & Perrachione, 2007).

With respect to talker variability, our results did not support Wang et al.'s (1999) finding that multitalker input leads to improved tonal categorization among classroom Mandarin L2 learners already familiar with tone. Unexpectedly, we observed no difference in overall 4-AFC accuracy between the single-talker and multitalker groups. A difference, however, was observed for the low-frequency syllables; for participants trained and tested on multiple speakers, low-frequency syllables with low-probability tones (F−P−) had the lowest average accuracy while low-frequency syllables with high-probability tones (F−P+) had the highest average accuracy (Figure 5B). For participants trained and tested on the same speaker, no such difference was found. There are at least three factors that may have contributed to this two-way interaction between syllable frequency and talker variability.

First, for true beginner learners like the participants tested in the present study, talker variability may have hindered learning the low-frequency, low-probability tones. There may have been too much variation in the F0 contours across the four talkers and too few

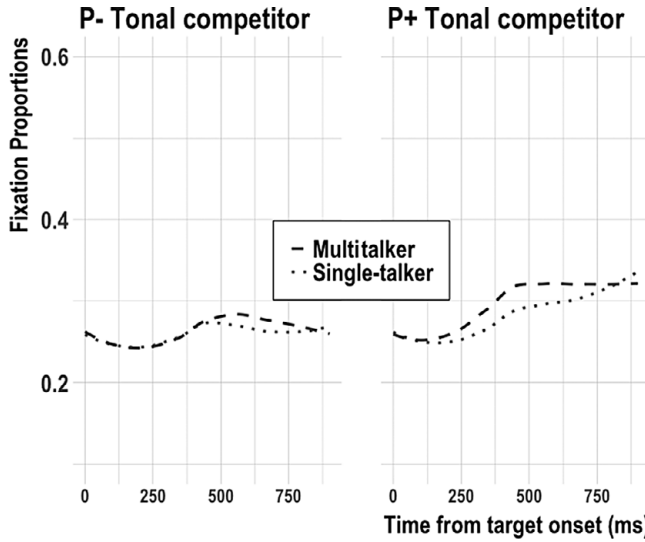

FIGURE 8. Day 4 *loess*-smoothed mean fixations to the tonal competitor across all high-probability P+ and low-probability P− trials by talker variability.

exemplars to learn from. Thus, the difference observed in 4-AFC results between the F− items may have reflected the challenge of learning highly varied tones with a limited number of exemplars. Under this account, as L2 learners accrue greater awareness of tonal categories and weigh F0 movement cues more heavily over time, F0 variability may prove to be more beneficial in tonal word learning (e.g., Barcroft & Sommers, 2014). For instance, intermediate and advanced learners may gradually encode more robust tone categories given more varied input, which is in line with Wang et al.'s (1999) results.

Second, greater variability in the input may have led to a greater tendency to rely on experience-based distributional processing as a means of accommodating the acoustic variability (e.g., Nixon & Best, 2018; Nixon et al., 2016; Wiener et al., 2018). This probability-based account is in line with the eye-fixation results. Figure 8 plots the *loess*-smoothed mean fixations to the low-probability (P−) and high-probability (P+) tonal competitors across all trials, irrespective of syllable frequency. This figure demonstrates that upon hearing the P+ target, participants looked to the P− tonal competitor at nearly identical rates across variability training conditions (Figure 8, left). In contrast, upon hearing the P− target (Figure 8, right), participants exposed to multiple talkers looked proportionally more often to the P+ tonal competitor from about 300 to 700 ms. This increase in looks to the more probable tonal competitor demonstrates an emerging tendency for learners to rely greater on probability-based tone processing in the face of multitalker input.

Further support for this claim was found in a post-hoc analysis[4] involving participants' overall 4-AFC rate of improvement from the first day of testing to the last day of testing (see Supplementary Materials for additional details including statistical analysis). This also acknowledges a third potential contributor to our results: individual differences in learning aptitude and/or perceptual abilities (among other factors; see Bowles et al., 2016;

Chandrasekaran et al., 2010; Theodore et al., 2019). Whereas none of the tested participants demonstrated a pitch deficiency, it remains possible that only participants with strong perceptual abilities benefited from the greater variability in the input while those participants with weak abilities saw no gains (e.g., Perrachione et al., 2011; Sadakata & McQueen, 2014). Our post-hoc analysis revealed that participants with lower four-day improvement rates relied more heavily on tonal probability information, and that this effect was particularly robust for those tested on multitalker input. In particular, listeners with lower improvement rates were less certain of the intended tonal category of the incoming speech signal and therefore relied more heavily on their knowledge of syllable-tone co-occurrence probabilities. The difference between the early eye-movement data and the resulting mouse clicks may have reflected learners' difficulties in recovering from their initial probability-based predictions, which led to the errors in the 4-AFC task (Figure 5B). In contrast, those learners with higher improvement rates may have been better at recovering from incorrect initial predictions and ultimately identifying the correct target. These preliminary results extend individual differences research (e.g., Perrachione et al., 2011; Sadakata & McQueen, 2014) by demonstrating how learners with lower tonal improvement rates may use statistical knowledge to bootstrap their learning of tone. We note, however, that recent evidence suggests different levels of learner aptitude may not necessarily lead to better or worse learning from high (or low) talker variability training input (Dong et al., 2019). Additional research in this domain is needed to clarify the role of individual differences in tone learning.

Finally, we acknowledge that the 4-AFC and eye-fixation results may also be accounted for by an error-driven learning mechanism (e.g., Nixon 2018, 2020) rather than a purely statistical learning mechanism. Similar to the present study, Nixon (2020) trained Native English speakers on an artificial tonal language and demonstrated that the cue-outcome order of stimuli affects tone learning. Nixon presented participants with a control cue (e.g., VOT) or a critical cue (e.g., tone) and trained participants to associate the cue with an outcome (e.g., image). After this pretraining phase, the critical cue was presented with a second cue (e.g., nasal) in the main training phase. Testing involved only the second cue. Nixon found that when the critical cue was learned in the pretraining phase, it "blocked" learning of the second cue (e.g., Kamin, 1969). Nixon next manipulated the temporal order of cues and outcomes by presenting either a discriminative order (cues followed by images) or nondiscriminative order (images followed by cues). Nixon found that discriminative structures, which allow for feedback from prediction error, resulted in better learning of predictive cues, particularly for items with fewer exemplars (i.e., low token frequency). Additionally, discriminative structures allowed better down-weighting or unlearning of nondiscriminative cues.

In the present study, we varied the cue-outcome order in different ways. The first two daily tasks (Figure 4) involved the simultaneous presentation of the syllable-tone (cue) and symbol (outcome). The naming task presented the nonce symbol (i.e., the outcome) first and required participants to produce the syllable-tone cue—a task analogous to presenting Chinese learners with a written character and asking them to produce its syllable-tone combination (see Chung, 2003). This may have allowed for feedback from prediction error in line with Nixon's (2020) design; however, all participants were given feedback irrespective of naming accuracy. It is unclear whether such feedback guided learning of incorrect responses (e.g., Arnon & Ramscar, 2012; Ramscar et al., 2010;

2013). Moreover, we note that feedback from our naming production task may work in a fundamentally different way than that from a perceptual identification task. For beginner L2 learners especially, production abilities may lag behind perceptual abilities, which could influence learning (e.g., Baese-Berk & Samuel, 2016). Our fourth and final daily task involved simultaneous presentation of the cue and outcome. For this 4-AFC identification task, feedback was again provided after each mouse click irrespective of accuracy. To what degree this feedback influenced the mouse-click and eye-fixation data remains unclear. Our future work will aim to test how the cue-order outcome and discriminative structures in line with Nixon's (2020) proposal contribute to the present study's observed learning of tone.

CONCLUSION

In this study we established that adult participants unfamiliar with lexical tone are able to track syllable conditioned tonal probabilities. This statistical learning occurs irrespective of the amount of talker variability in the speech signal or participants' explicit instructed awareness of tone. Future research is needed to fully tease apart whether the underlying mechanism driving this learning is statistical in nature (e.g., Maye et al., 2002; Saffran, 2003) or error driven (e.g., Nixon, 2020) and to what degree individual differences affect this learning. Finally, there is a need for future research to examine the pedagogical implications of this work and investigate how classroom Mandarin learners track the occurrence and co-occurrence regularities of syllables and tones and to what degree the timing of acoustic cues and semantic outcomes affect this learning.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0272263120000418.

NOTES

[1]All participants were able to reliably discriminate two pure tones differing by 20 Hz (or lower) as a screening threshold. Due to a data logging error, however, individual results were lost and thus not included as covariates in the regression analysis.

[2]We acknowledge that the four tasks are not independent of one another. See Wiener (2015) for a complete discussion of the tasks, which is beyond the scope (and word limit) of the present study.

[3]Reported *p*-values obtained from releveling the model.

[4]We thank the anonymous reviewer for suggesting this post-hoc analysis.

REFERENCES

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*, 292–305.

Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36.

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

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.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* (Version 6.0.19) [Computer program]. http://www.praat.org

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, *66*, 774–808.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLOS ONE*, *5*, e10729.

Caldwell-Harris, C. L., Lancaster, A., Ladd, D. R., Dediu, D., & Christiansen, M. H. (2015). Factors influencing sensitivity to lexical tone in an artificial language. *Studies in Second Language Acquisition*, *37*, 335–357.

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.

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*, 3703–3716.

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.

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.

Chen, J. Y., Lin, W. C., & Ferrand, L. (2003). Masked priming of the syllable in Mandarin Chinese speech production. *Chinese Journal of Psychology*, *45*, 107–120.

Chun, D. M., Jiang, Y., Meyr, J., & Yang, R. (2015). Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. *Journal of Second Language Pronunciation*, *1*, 86–114.

Chung, K. K. (2003). Effects of Pinyin and first language words in learning of Chinese characters as a second language. *Journal of Behavioral Education*, *12*, 207–223.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*, 829–836.

Colantoni, L., Steele, J., Escudero, P., & Neyra, P. R. E. (2015). *Second language speech*. Cambridge University Press.

Cutler, A. (2012). *Native listening*. MIT Press.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*, 317–367.

DeFrancis, J. (1986). *The Chinese language: Fact and fantasy*. University of Hawaii Press.

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.

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

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

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188.

Ellis, N. C. (2011). Frequency-based accounts of SLA. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 193–210). Routledge/Taylor Francis.

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*, EL206–EL212.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*, 551–585.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*, 345–360.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. (2016). The relationship between artificial and second language learning. *Cognitive Science*, *40*, 822–847.

Fox, R. A., & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics*, *13*, 69–90.

Gandour, J. (1983). Tone perception in far eastern-languages. *Journal of Phonetics*, *11*, 149–175.

Godfroid, A., Lin, C. H., & Ryu, C. (2017). Hearing and seeing tone through color: An efficacy study of web-based, multimodal Chinese tone perception training. *Language Learning*, *67*, 819–857.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*, 178–186.

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.

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

Hao, Y. C. (2018). Second language perception of Mandarin vowels and tones. *Language and Speech*, *61*, 135–152.

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

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

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

Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, *33*, 353–367.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*, 3059–3071.

Howie, J. (1976). *Acoustical studies of Mandarin vowels and tones*. 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.

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.

Idemaru, K., & Holt, L. L. (2020). Generalization of dimension-based statistical learning. *Attention, Perception, & Psychophysics*, *82*, 1744–1762.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment aversive behavior* (pp. 279−296). Appleton-Century-Crofts.

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

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

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*, 864–886.

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

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

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

Li, P., & Yip, M. C. (1998). Context effects and the processing of spoken homophones. In C. K. Leong & K. Tamaoka (Eds.), *Cognitive processing of the Chinese and the Japanese languages* (pp. 69–89). Kluwer Academic Publishers.

Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a tonal language by attending to the tone: An in vivo experiment. *Language Learning*, *61*, 1119–1141.

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*, 1242–1255.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *64*, 407–420.

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, *50*, 2032–2043.

Mandell, J. (2015). *Tonometric* [Computer software]. http://jakemandell.com/adaptivepitch/

Matin, E., Shao, K.C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Attention, Perception, & Psychophysics*, *53*, 372–380.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.

McQueen, J. M., & Cutler, A. (2010). Cognitive processes in speech perception. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.) *The handbook of phonetic sciences* (Vol. 1) (pp. 489–520). Blackwell Publishing.

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.

Myers, J. (2002). An analogical approach to the Mandarin syllabary. *Journal of Chinese Phonology*, *11*, 163–190.

Myers, J. (2010). Chinese as a natural experiment. *The Mental Lexicon*, *5*, 423–437.

Nixon, J. S. (2018). Effective acoustic cue learning is not just statistical, it is discriminative. In *Proceedings of INTERSPEECH* (pp. 1447–1451).

Nixon, J. S. (2020). Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition*, *197*, e104081. https://doi.org/10.1016/j.cognition.2019.104081.

Nixon, J. S., & Best, C. T. (2018). Acoustic cue variability affects eye movement behaviour during non-native speech perception. In *Proceedings of the 9th International Conference on Speech Prosody* (pp. 493–497).

Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: Eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language*, *90*, 103–125.

Ong, J. H., Burnham, D., & Escudero, P. (2015). *Distributional learning of lexical tones: A comparison of attended vs. unattended listening*. *PLOS ONE*, *10*, e0133446.

Ong, J. H., Burnham, D., Escudero, P., & Stevens, C. J. (2017). Effect of linguistic and musical experience on distributional learning of nonnative lexical tones. *Journal of Speech, Language, and Hearing Research*, *60*, 2769–2780.

Packard, J. L. (1999). Lexical access in Chinese speech comprehension and production. *Brain and Language*, *68*, 89–94.

Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press.

Pajak, B., Fine, A. B., Kleinschmidt, D. F., & Jaeger, T. F. (2016). Learning additional languages as hierarchical probabilistic inference: insights from first language processing. *Language Learning*, *66*, 900–944.

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*, 51–78.

Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2019). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition*, *41*, 59–86.

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.

Potter, C. E., Wang, T., & Saffran, J. R. (2017). Second language experience facilitates statistical learning of novel linguistic materials. *Cognitive Science*, *41*, 913–927.

Qin, Z., Tremblay, A., & Zhang, J. (2019). Influence of within-category tonal information in the recognition of Mandarin-Chinese words by native and non-native listeners: An eye-tracking study. *Journal of Phonetics*, *73*, 144–157.

Qin, Z., & Zhang, C. (2019). The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PLOS ONE*, *14*, e0221498.

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of "mouses" in adult speech. *Language*, 760–793.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*, 909–957.

R Core Team. (2019). *R: A language and environment for statistical computing* [Computer program]. Version 3.6.1. http://www.r-project.org

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*, 1318.

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110–114.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101–105.

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language Awareness*, *20*, 45–59.

Saito, K. (2015). Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English/ɹ/without explicit instruction. *Applied Psycholinguistics*, *36*, 377–409.

Saito, K., & Wu, X. (2014). Communicative focus on form and second language suprasegmental learning. *Studies in Second Language Acquisition*, *36*, 647–680.

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

Shen, X. S., & Lin, M. C. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech*, *34*, 145–156.

Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research*, *29*, 185–200.

Theodore, R. M., Monto, N. R., & Graham, S. (2019). Individual differences in distributional learning for speech: What's ideal for ideal observers? *Journal of Speech, Language, and Hearing Research*, 1–13.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*, 11177–11187.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*, 325–329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374–408.

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.

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*, *64*, 122–144.

Wang, H. S. (1998). An experimental study on the phonotactic constraints of Mandarin Chinese. *Studia Linguistica Serica*, 259–268.

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, *5*, 953.

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

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.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*, 1–25.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*, *119*, 597–607.

Wewalaarachchi, T. D., Wong, L. H., & Singh, L. (2017). Vowels, consonants, and lexical tones: Sensitivity to phonological variation in monolingual Mandarin and bilingual English–Mandarin toddlers. *Journal of Experimental Child Psychology*, *159*, 16–33.

Wiener, S. (2015). *The representation, organization and access of lexical tone by native and non-native Mandarin speakers* (Unpublished doctoral dissertation). The Ohio State University.

Wiener, S., Chan, M. K. M., & Ito, K. (2020). Do explicit instruction and high variability phonetic training improve non-native speakers' Mandarin tone productions? *The Modern Language Journal*, *104*, 152–168.

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.

Wiener, S., & Ito, K. (2016). Impoverished acoustic input triggers probability-based tone processing in monodialectal Mandarin listeners. *Journal of Phonetics*, *56*, 38–51.

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.

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.

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*, 527–558.

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

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

Xing, J. Z. (2006). *Teaching and learning Chinese as a foreign language: A pedagogical grammar* (Vol. 1). Hong Kong University Press.

Yang, B. (2015). *Perception and production of Mandarin tones by native speakers and L2 learners*. Springer Publisher.

Yin, B. Y. (1984). Hanyu yusu de dingliang yanjiu [A quantitative research of Chinese morphemes]. *Zhongguo Yuwen [Chinese Literature and Language]*, *182*, 338–347.

Zee, Y.-Y. (1980). *Phonetic studies of Chinese tones* (Unpublished doctoral dissertation). University of California.

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

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

Zipf, G. K. (1935). *The psychobiology of language*. Houghton Mifflin.

Zipf, G. K. (1949). *Human behavior and the principle of least-effort*. Addison-Wesley.