

RESEARCH ARTICLE



Effects of high accent variability on monolingual English perceivers' acquisition of Mandarin minimal-tone-contrast words

Yanping Li¹ , Steffen A. Herff^{1,2}, Michael D. Tyler^{1,3}, Denis Burnham¹ and Catherine T. Best¹

¹The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia; ²Sydney Conservatorium of Music, University of Sydney, Sydney, Australia and ³School of Psychology, Western Sydney University, Sydney, Australia

Corresponding author: Yanping Li; Email: yanping.li@westernsydney.edu.au

(Received 16 May 2025; Revised 28 February 2026; Accepted 30 March 2026)

Abstract

Natural phonetic variability such as talker differences facilitates second language (L2) word learning. Whether accent variability enhances the talker variability benefit, particularly in the learning of words that differ only in lexical tones, has not yet been examined despite its potential applied value for L2 tone language learning. Two groups of monolingual English speakers completed six training sessions on four minimal-tone quadrads of Mandarin monosyllabic pseudowords, produced either by 12 talkers from Beijing (single accent) or by four talkers each from Beijing, Yantai, and Guangzhou (multiple accents). Bayesian mixed-effects modeling revealed strong learning improvement across sessions in both conditions, but the multiple-accent group improved faster than the single-accent group. In addition, the multiple-accent group demonstrated superior generalization to new talkers with a familiar and an unfamiliar accent, suggesting that natural L2-Mandarin accent variability facilitates English learners' access to abstract tone-word lexical representations, beyond the benefits of talker variability alone.

Keywords: lexical tone contrasts; high variability training; Mandarin words; L2 accents; non-tone learners

Introduction

Non-tone-language (e.g., English) learners of a tone language as an L2 (second language) experience difficulties in distinguishing between words that only differ in lexical tone (Pelzl et al., 2021). Lexical tones are based on, and contrasted by, the height and contour of their fundamental frequency (F0, perceived as pitch). For example, the consonant-vowel [CV] syllable /ma/ in Mandarin Chinese has four different meanings when carrying each of the four Mandarin Chinese tones: level tone = *mother*, rising = *hemp*, dipping = *horse*, and falling = *scold*. Although talker variability during training appears to facilitate L2 tone-word learning (Dong et al., 2019; Perrachione et al., 2011;

see also Uchihara *et al.*, 2025, for a meta-analysis of high-variability phonetic training), the effects of other types of natural variability, such as regional accent, on L2 lexical tone learning remain unexplored. As a higher-order level of phonetic variability, regional accent differences further modulate talker-specific phonetic patterns in speech. Both talker and accent variability are ubiquitous in natural spoken language, and increased variability has been argued to promote more abstract and robust phonological representations. If talker variability in the input benefits tone-language learning, then higher-order accent variability may further enhance this effect, potentially improving L2 word learning. While the effects of phonetic variability (e.g., talker and/or regional variability) on L2 word learning have been investigated (Barcroft & Sommers, 2014; Fichtner *et al.*, 2025; Uchihara *et al.*, 2022), these studies have focused almost exclusively on non-tone languages. Given that over half of the world's population speaks a tone language as either a first language (L1) or L2 (Fromkin, 1978), regional and foreign accent variability is widespread across tone languages (Li *et al.*, 2020a). Accordingly, the present study investigates the effects of high accent variability in lexical tones on Mandarin tone-word learning by novice English learners.

Lexical processing of tone contrasts and L2 tone-word training

Non-tone-language learners of tone languages are capable of categorically identifying and discriminating tone categories (Hao, 2012, 2018), and naïve listeners can achieve a high degree of mastery in identifying and discriminating them through focused training (Chandrasekaran *et al.*, 2010). However, excellence in tone identification and discrimination alone does not necessarily translate to improved *lexical* processing of words differentiated by tone contrasts (Pelzl *et al.*, 2021). Difficulty with tone distinctions in their lexical contexts may stem from the linguistic status of F0 variation within the phonological hierarchy of learners' L1s (Best, 2019). As shown in Table 1, all languages use F0 variation linguistically to mark intonation distinctions at the supra-lexical level (sentences, phrases) and the lexical level (lexical emphasis/focus, stress or pitch accent contrasts) of the phonological hierarchy (Nespor & Vogel, 2007). In addition, tone languages use not only consonant and vowel segmental distinctions but also F0 distinctions (analogous to segments) at the sub-lexical level, as phonological sub-components of words. That is, they use contrastive F0 specifications at all three, rather than only two, levels of the phonological hierarchy. On the other hand, non-tone languages such as English have a phonological gap in contrastive use of F0 at the sub-lexical (segmental) level. This gap may explain non-tone-language listeners' persistent difficulties in perceiving and remembering the lexical tones of words in L2 tone languages (Pelzl *et al.*, 2021), even though they can discriminate and learn to categorize/label the tones fairly well as isolated pitch patterns (Burnham *et al.*, 2015; Hao, 2012). It may thus be necessary for lexical tone training of English tone-L2 learners to promote the perception of F0 as a sub-lexical segmental phonological feature, rather than as a supra-lexical/lexical phonological pattern or as a paralinguistic property of the stimulus (Liu *et al.*, 2022; So & Best, 2014).

Table 1. The F0 variation used in tone and non-tone languages at different levels of the phonological hierarchy

Levels of phonological hierarchy	Tone languages	Non-tone languages
Supra-lexical (phrasal intonation)	✓	✓
Lexical (emphasis/focus, stress/pitch accent)	✓	✓
Sub-lexical (segmental-level distinctions)	✓	✗

Indeed, training L2 non-tone-language learners to acquire tone-contrastive words does appear to confer a benefit on being able to use tones in a lexical context. For example, Wong and Perrachione (2007) used a sound-to-meaning paradigm to train novice English learners in tone acquisition through lexical identification. Participants were asked to associate tonal pseudowords with drawings representing high-frequency English nouns. The pseudowords were not real words in either English or Mandarin (or other tone languages), and the meanings associated with the drawings were arbitrarily assigned. Successful learners improved tone-word identification and reached asymptotic performance after seven training sessions, suggesting that tone training in a lexical context fostered the use of pitch variation as an integral phonological feature of the words, a conclusion further supported by recent L2 word training studies (Ge et al., 2026), including tone-word learning studies (Laméris et al., 2024). It could be inferred that linking tone contours to lexical meanings can help non-tone-language novice learners establish robust phonological tone categories at the sub-lexical segmental level. These findings suggest that L2 speech training, particularly for tone acquisition, may be most effective when embedded in a lexical context. If lexical-context training facilitates the establishment of tone categories, an important next question concerns the conditions under which these emerging representations become robust and generalizable. The following section reviews this issue.

High-talker variability training and phonological constancy in L2 word learning

High-talker variability phonetic training has been shown to facilitate L2 speech perception (Uchihara et al., 2025). More specifically, variability in speaking styles (speech registers), speaking rates, and talkers introduced during training has been found to benefit L2 word learning. For example, novice English learners of Spanish achieved higher word-recall accuracy when exposed to concrete nouns spoken by six talkers (high-talker variability) than when exposed to only three talkers or just a single talker (low-talker variability; Barcroft & Sommers, 2005). These findings suggest that exposure to high-talker variability enhances L2 word learning, potentially by encouraging learners to extract invariant phonological representations across variable input. However, in these studies, generalization tests to new speaking styles, speaking rates, or talkers were not included, leaving it unclear whether learners had established robust, abstract phonological representations. Studies on L2 tone-word learning have addressed this concern by incorporating generalization tests with new talkers, for example, allowing researchers to directly assess whether learners have established abstract tone representations. For instance, following L2 tone-word training, novice English learners were tested on the trained words spoken by new talkers. Participants in the high-talker variability group outperformed those in the low-talker variability group on word identification in those post-training tests, indicating superior talker-general generalization (Perrachione et al., 2011). High-talker variability during training also facilitated the weighting of lexical tone cues, shifting reliance from pitch height to pitch contour (Chandrasekaran et al., 2010; Sadakata & McQueen, 2014), the latter being the primary cue used by tone-language L1 listeners for tone-specific word identification (Gandour, 1983). These findings indicate that high-talker variability promotes the development of robust tonal representations in a lexical context, providing empirical support for its role in L2 tone-word learning.

However, although the prior studies suggest that high-talker variability can facilitate L2 tone-word learning, some findings indicate that such variability does not consistently confer an advantage in L2 speech learning. For example, Dong et al. (2019) examined Mandarin tone-word training for novice English learners under conditions

of high-talker versus low-talker variability and observed comparable learning and generalization outcomes between the two groups. This discrepancy may reflect differences in the training materials and task demands. Many earlier studies used synthesized or hybrid stimuli, resynthesizing English syllables produced by L1-English talkers with lexical tone contours produced by L1-Chinese talkers (Chandrasekaran *et al.*, 2010), whereas Dong *et al.* (2019) used *real* Mandarin words encompassing all *six* Mandarin vowels (cf. Sadakata & McQueen, 2014, who used five, excluding /y/) combined with all *four* tones (cf. Perrachione *et al.*, 2011, who used only three, excluding the dipping tone), increasing phonetic and lexical complexity relative to previous work. As the authors noted, this higher cognitive load may have attenuated the potential benefits of phonetic variability for L2 tone-word learning. Accordingly, effective L2 tone-word training should maintain ecologically valid phonetic variability while reducing unnecessary cognitive demands, thereby creating conditions under which variability can promote robust phonological abstraction.

The empirical benefits of high-talker variability under such conditions raise a key question: what perceptual and cognitive mechanisms enable exposure to multiple talkers to foster robust, generalizable lexical tone representations? The answer may be found in considering the fact that natural exemplars of a given spoken word display substantial but lexically irrelevant phonetic variability across talkers and their emotional states, speaking styles, and other indexical factors that reflect properties of the talker rather than the linguistic content of the message. The Perceptual Attunement account (Best, 2015) suggests that acquiring and recognizing words requires a learner to distinguish which phonetic differences signal a change in meaning (*phonological distinctiveness*) and which other phonetic differences, such as talker-specific indexical information, leave the identity of the word intact (*phonological constancy*). Crucially, phonological constancy depends on phonological abstraction, a core assumption of abstractionist models of spoken word recognition (McQueen *et al.*, 2006), which posit that lexical representations encode linguistically relevant phonological structure while abstracting away from indexical acoustic variation associated with individual talkers and exemplars. Consequently, purely exemplar-based representations (Johnson, 1997) are insufficient to support phonological constancy, because they can only support recognition of words produced in previously experienced voices. Hybrid models of spoken word representation (Cutler & Weber, 2007; Goldinger, 2007) reconcile these positions by hypothesizing that perceivers (including learners) retain exemplar-specific detail while organizing it around abstract phonological representations, thereby enabling both sensitivity to indexical variability and generalization to new talkers. For L2 learners, indexical variability in words may hinder lexical acquisition unless phonological constancy has emerged. Thus, incorporating indexical variability during L2 word learning may foster abstraction and support the development of robust L2 lexical representations.

High-variability training across multiple talkers exposes learners to talker-specific phonetic differences while preserving word identity, but there are other types of indexical variability that reflect broader, shared phonetic patterns across groups of talkers. It is reasonable to predict that presenting learners with indexical variability beyond individual talkers may further support the extraction of stable phonological representations and promote L2 lexical learning. Importantly, this should be reflected in more robust generalization to a new group of talkers. Regional accent differences are one such type of group-level indexical variability. The next section will review studies that incorporate accent variability to investigate its effects on L2 tone-word learning.

Incorporating accent variability to L2 (tone) word learning

Both talker and accent variation generate idiosyncratic and systematic acoustic-phonetic differences in words, respectively, constituting critical sources of variability that L2 learners must learn to accommodate. Talker variation gives rise to idiosyncratic differences in the realization of lexical tones across individual talkers (between-talker variation; Tyler & Best, *in preparation*; see also Foulkes & Hay, 2015). These differences may stem from anatomical and physiological properties of the vocal apparatus (e.g., F0 dynamic range, voice quality) as well as from indexical characteristics of individual talkers that reflect broader group memberships, such as gender, age, and sociocultural factors (e.g., education level, social class). Beyond individual talkers, accent variation reflects broader, shared phonetic patterns across groups of talkers, introducing an additional and more abstract source of systematic acoustic-phonetic variability (between-group variation). Together, these sources of variability shape how L2 learners extract stable lexical representations from variable speech input.

Accent variation encompasses systematic differences in speech production that arise from shared phonetic patterns across groups of talkers. Such variation may reflect regional accent differences among L1 talkers of the target language (typical English), as in British versus American varieties, or systematic deviations in L2 talkers' productions relative to those of L1 talkers. For example, the words *rock* and *lock* are realized as /ɹɔk/ and /lɔk/ (mid-low back rounded vowel) in British English, but as /ɹɛk/ and /lɛk/ (low mid-central vowel) in American English. These same words may also be produced as [roku] by Japanese learners of L2 English, reflecting the influence of Japanese phonology, in which the English liquids /l/ and /ɹ/ are both commonly realized as the alveolar tap /ɾ/ in syllable-initial position (Tyler, 2021), a high back unrounded vowel is appended to the final consonant, and the stressed vowel is mid-high back rounded. These examples illustrate how accent-related variation introduces systematic acoustic-phonetic differences that extend beyond individual talker idiosyncrasies and must be accommodated in L2 lexical learning. The present study focused on training effects induced by accent variability in addition to talker variability, both of which constitute critical sources of indexical information in speech (Best, 2015). Since the usually beneficial effects of talker variability have already been shown in prior research, talker variability was held constant across low-accent and high-accent variability training conditions in the current design.

Exposure to regional accents has been shown to facilitate L2 word learning. For example, Fichtner et al. (2025) incorporated accent variability into L2 word learning by training novice English learners on German words spoken by six talkers either in a single accent (multiple talkers, single accent) or in six different varieties of German (multiple talkers, multiple accents). Learners were tested on word recall, not identification of the learned words, so the study did not include tests of generalization to new talkers or accents. More words were recalled in the multiple-German-accent condition than in the single-accent condition, an effect that held when word groups and learning conditions were counterbalanced. To the best of our knowledge, Fichtner et al. (2025) provide the most directly relevant evidence on the effects of accent variability in L2 word learning. However, neither English nor German is a tone language, so the findings of Fichtner et al. cannot shed light on whether regional accent variability could enhance L2 tone-word learning by learners who lack lexical tones in their L1s (i.e., a phonological gap at the sub-lexical level; see Table 1). Therefore, the present study investigated the effect of regional accent variability versus talker variability alone on novel tone-word learning by non-tone-language speakers. We hypothesized that learners exposed

to regional accent variability would learn tone-words better than those exposed to talker variability alone. Crucially, we hypothesized that this learning would generalize more effectively to new talkers with a familiar and an unfamiliar regional accent than in the single-accent training condition.

The present study

The learners in this study were native English speakers from non-tone-language backgrounds. For the stimulus tone language, we chose Mandarin Chinese, as it is widely spoken, widely geographically distributed, and has extensive regional accent variability. Mandarin Chinese shows substantial dialectal variation in mainland China, represented across 7–10 distinct dialect groups that include the *Mandarin*, *Wu*, and *Yue* dialect groups, which are regarded as different languages and are often not mutually intelligible. Beijing dialect is the prestige variety of the *Mandarin* group, which is designated as Standard Mandarin (henceforth, Mandarin). Children raised in Beijing acquire Mandarin as their L1 and their pronunciation is the standard accent of Mandarin. Children from other regions of mainland China acquire other regional Chinese languages as their L1s, which are widely spoken at home and more generally within the home region. They start to learn Mandarin as their L2 upon enrolment in primary school (i.e., from around six years of age). After years of immersion and daily use of Mandarin, most Chinese adults residing outside Beijing speak Mandarin fluently. Therefore, they are fluent early-onset L2-Mandarin talkers.

Regional talkers' L2 pronunciation of Mandarin tones can deviate from native Mandarin norms due to influences of their L1s, which often have different tone systems from Beijing Mandarin and/or different phonetic realizations of shared tones. Yantai, Shanghai, and Guangzhou are different from each other, as well as from Mandarin (see Li *et al.*, 2020b for more details about their differences in tone inventories), with differing tone inventories among them. Li *et al.* (2020a, 2020b) examined their productions of Mandarin tones in detail and found that while L2-Mandarin talkers from each region clearly distinguish the four tones, their productions of each tone deviate acoustically (phonetically) from L1-Mandarin, resulting in regionally accented Mandarin lexical tones. As shown in Figure 1, the F0 contours of the Mandarin level tone for L1-Mandarin and the three L2-Mandarin accents are all straight lines with high F0, but there are subtle rising or falling differences at some parts of the level tone's F0 contours. Mandarin rising and falling tones show a trajectory of rising and falling, respectively, in all four accents, but the steepness of the slopes differs across accents. F0 contours of L2-Mandarin dipping tones in the Shanghai (long-dash line in Figure 1) and Guangzhou (dotted line) accents are shallower than those in L1-Mandarin (solid line) and Yantai L2-Mandarin (dot-dashed line). Moreover, the dipping curve in the Guangzhou accent is much shallower than that of the Shanghai accent, yielding an overall higher F0 height in Guangzhou. It is these F0 deviations that are expected to facilitate English learners' tone-word learning and generalization to new talkers with a familiar and an unfamiliar accent in a high-variability training task that includes both accent and talker variability, as compared to those trained only with talker variability.

We extended previous research on L2 tone-word training with talker variability, by training native English speakers on Mandarin minimal-tone contrasts in a high-talker-variability tone-word learning task with either the Beijing accent alone (single

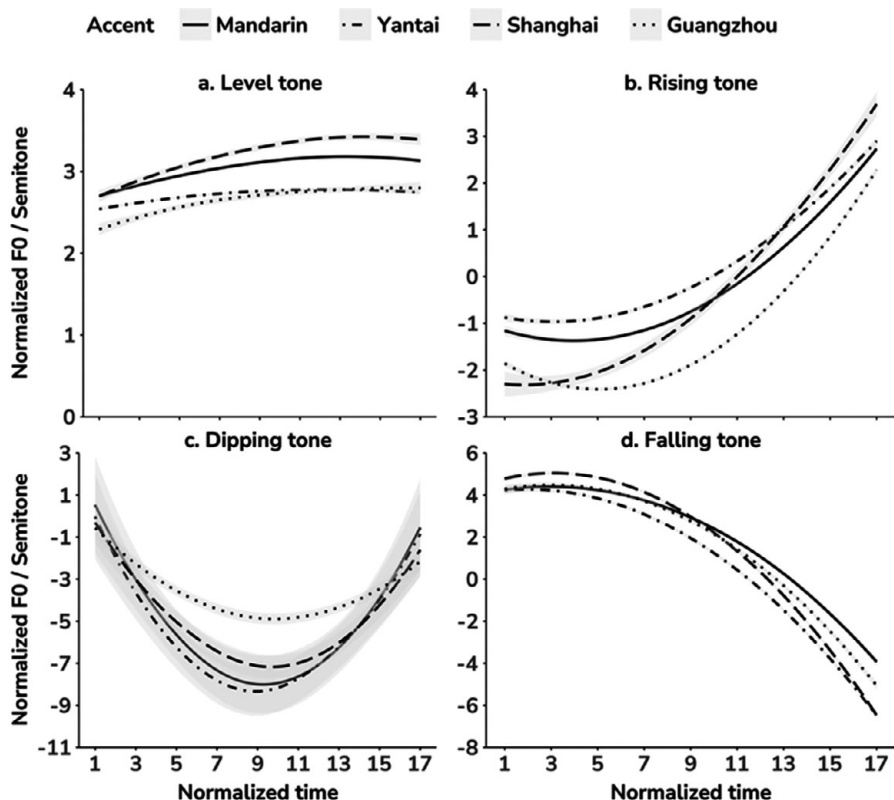


Figure 1. F0 variations of the Mandarin tones in L2-Mandarin accents in relation to the L1-Mandarin accent for the stimuli used in this study. Lines indicate the mean normalized F0 contours across four talkers of each accent in normalized time. The stimuli and this figure are adapted from Li et al. (2020b).

accent) or with the Beijing, Yantai, and Guangzhou accents (multiple accents). We predicted:

- 1) Learners trained on words spoken in the Beijing, Yantai, and Guangzhou accents will achieve better tone-word identification than learners trained on words spoken in the Beijing accent only.
- 2) In the test of generalization to new talkers, learners trained on words spoken in Beijing, Yantai, and Guangzhou accents will identify tone-words more accurately than learners trained on the Beijing accent only, for both the familiar Beijing accent and especially for the unfamiliar Shanghai accent.

Materials and methods

Participants

Forty-eight L1-English speaking students were recruited online from Australian universities and randomly assigned to a single-accent ($n = 24$, $M_{\text{age}} = 24.5$ years, $SD = 5.8$, 14 females) or multiple-accent ($n = 24$, $M_{\text{age}} = 25.5$ years, $SD = 5.1$, 15 females) group. All had acquired English from birth, either in Australia ($n = 46$, 27 females) or in the

United States ($n = 2$, 2 females). Some of them had learned other language(s) such as Spanish and/or French, but all were linguistically naïve to Mandarin or any other lexical tone languages. Because sensitivity to pitch height and contour (i.e., pitch aptitude) is necessary for the adequate perception and learning of lexical tones (Perrachione *et al.*, 2011; Wong & Perrachione, 2007), all participants demonstrated high pitch aptitude, as evidenced by above-chance performance on a pre-training F0 height and contour discrimination test (see [Supplementary Materials](#) for details: <https://doi.org/10.17605/OSF.IO/C6SBM>). Given evidence that extensive musical training facilitates non-native tone perception and L2 tone-word learning (Laméris *et al.*, 2024), only non-musicians were recruited (defined as having no more than three cumulative years of private lessons across any combination of instruments, Wong & Perrachione, 2007). Their language and music backgrounds were self-reported through an online survey, which also confirmed that none of them had any speech or hearing disorders.

This study was approved by the Western Sydney University Human Research Ethics Committee (Ethics Reference ID: H13105). Participants signed informed consent prior to testing and received eGift smart cards or course credits for their participation.

Stimuli

For the training and word identification tests, we selected 16 Mandarin words comprised of easily distinguished stop consonant-vowel (CV) syllables /ba¹/, /di/, /du/, and /gu/, combined with all four Mandarin tones (see [Table 2](#)). We included the dipping tone, which had been previously excluded in Perrachione *et al.* (2011) due to its confusion with the rising tone (Hao, 2012). Our target lexicon instead reflected the full Mandarin tone system across multiple CV syllables, and thus was more challenging to learn than the 3-tone set used in Perrachione *et al.* (2011).

The single-accent training group (talker-only variability) were trained with multiple tokens of all target words by 12 L1 female talkers of Beijing Mandarin ($M_{\text{age}} = 21.83$ years, $SD = 1.99$). The multiple-accent training group (talker and accent variability) were also trained with multiple tokens of all words by 12 female talkers ($M_{\text{age}} = 21.58$ years, $SD = 2.19$), but the talkers were four L1 Mandarin talkers from Beijing ($M_{\text{age}} = 23.50$ years, $SD = 1.00$) and four early-fluent L2-Mandarin talkers each from Yantai ($M_{\text{age}} = 21.25$ years, $SD = 1.50$) and Guangzhou ($M_{\text{age}} = 19.75$ years, $SD = 1.26$). Both training groups included recordings from 12 talkers, representing high-talker

Table 2. Monosyllabic Mandarin words used in tone-word training and identification tasks.

Mono syllables	Mandarin tones			
	1—Level tone	2—Rising tone	3—Dipping tone	4—Falling tone
ba	疤 ba1	拔 ba2	把 ba3	坝 ba4
di	低 di1	笛 di2	底 di3	弟 di4
du	督 du1	读 du2	赌 du3	肚 du4
gu	姑 gu1	鹈 gu2	骨 gu3	固 gu4

¹/di/-/du/ differ in their vowels, and /du/-/gu/ differ in their initial consonants, forming minimal pairs. In contrast, /ba/ does not form a minimal pair with the other three syllables. Due to phonotactic constraints on Mandarin minimal quadrad words, only these four syllables were used. Importantly, this constraint did not confer any significant learning advantage for the /ba/ quadrads.

variability in both groups; they were differentiated only by the level of *accent* variability (one vs. four accents).

We recorded the target word materials from the native talkers under controlled conditions in a sound-attenuated booth at the Speech Acquisition and Intelligent Technology Lab, Beijing Language and Cultural University, China. Their productions were captured by a Shure SM57 dynamic microphone and recorded in Audacity (v. 2.3.0) on a Dell desktop computer at a 44.1-kHz sampling rate (32-bit resolution) via an M-audio M-track II external sound card. The Chinese characters displayed in Table 2 were presented individually per trial on an external monitor via E-Prime Professional 2.0 (Psychology Software Tools, Inc.) running on a Dell Latitude 7280 laptop. Each talker recorded 10 tokens of each word, once each per 10 recording blocks, yielding 160 trials per talker (16 words $\times$ 10 blocked repetitions), with the 16 words within each block presented in randomized order. Recordings were high-pass filtered at 70 Hz and manually segmented into individual utterances using Praat (Boersma & Weenink, 2018). The resulting stimuli were subsequently used in an accent-similarity test, administered in E-Prime on a Dell OptiPlex 3030 AIO desktop computer and presented over AKG K272 headphones. Accent similarity was verified by four new Beijing L1-Mandarin listeners for the Beijing accent, and by four new Yantai and four new Guangzhou L2-Mandarin listeners for the Yantai and Guangzhou accents, respectively. Listeners rated how similar each Mandarin-word pronunciation was to that of their native region on a 7-point Likert scale (1 = not at all similar, 7 = highly similar). For each talker and training condition, the four tokens of each word receiving the highest similarity ratings were selected for use in the training study, resulting in a total of 768 stimuli (12 talkers $\times$ 16 pseudowords $\times$ 4 tokens).

For investigating the generalization of the trained words to new talkers of the familiar Beijing accent and the unfamiliar (untrained) Shanghai accent, multiple tokens of the 16 Mandarin words in Table 2 were recorded by two additional female talkers as described above, one a 13th L1 talker of Beijing Mandarin (19.00 years; familiar accent) and the other an early fluent L2-Mandarin talker from Shanghai (24.00 years; unfamiliar accent). Neither talker was presented to the participants during training before the generalization tests, i.e., both were new talkers. As before, four tokens per word by each of the new talkers were selected for the post-training generalization tests, yielding 64 tokens per talker (16 words $\times$ 4 tokens). As with the training stimuli, the selected tokens were verified by four new Beijing L1-Mandarin or four new Shanghai L2-Mandarin female talkers.

Since our task required easily recognizable objects, the assigned English meanings for the 16 words were arbitrary, resulting in Mandarin pseudowords. The meanings (see Figure 2), which referred to depictable concrete objects with moderate to high word frequency, were selected from van Heuven et al. (2014). They were depicted with gray-scaled drawings with high naming agreement and low visual complexity selected from the Multilingual Picture (MultiPic) database (Duñabeitia et al., 2018). Word-meaning pairings were counterbalanced across participants (e.g., the cow corresponded to different Mandarin words such as 疤 /ba1/ and 拔 /ba2/ for different participants).

Procedure

The participants completed six training sessions on the 16 pseudowords over six consecutive days with a sound-to-meaning paradigm adapted from Wong and Perrachione (2007). Each training session presented four talkers of the same accent (approximately 45 min), yielding three sessions for a total of 12 talkers. Set order was counterbalanced

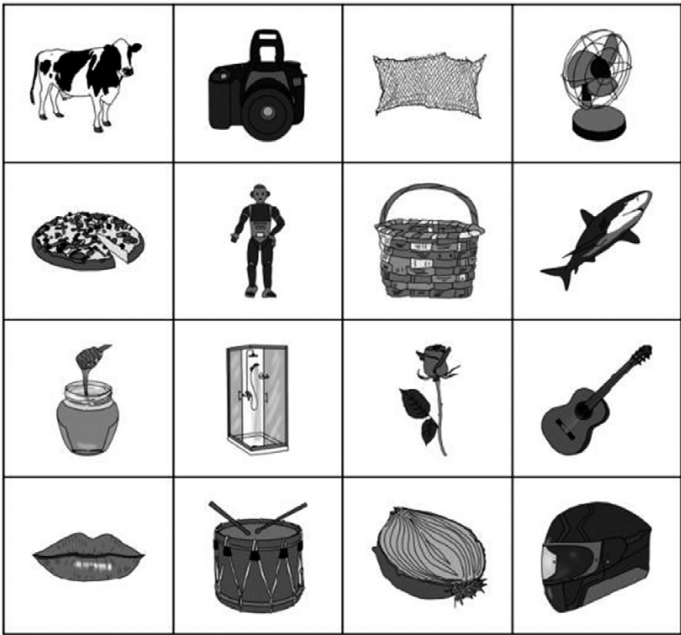


Figure 2. The 16 drawings for training.

across participants, and those three sessions (Sessions 1–3: Block 1) were repeated a second time (Sessions 4–6: Block 2) for a total of six sessions to ensure sufficient training as recommended in Wong and Perrachione (2007). The 12 Beijing Mandarin talkers for the single-accent group were randomly assigned to the three sessions per block, and the four talkers each of the Beijing, Yantai, and Guangzhou accents in the multiple-accent group were blocked by session. There were 256 randomized trials (4 talkers $\times$ 4 syllables $\times$ 4 tones $\times$ 4 tokens) in each session with 16 words blocked by talker ($n = 4$) to optimize word learning (Perrachione et al., 2011). For the same reason, talker blocks were further subdivided by syllable (i.e., by minimal-tone-contrast word set), yielding 16 randomized trials per word set (4 words $\times$ 4 tokens). Talker and syllable orders within talker were randomized across participants.

On each tone-word training trial, the target object image was displayed, starting 0.5 s earlier than the paired word began; the image remained on screen during the word presentation. Learners were instructed to remember as many word-meaning pairs as possible. After training on the four minimal-tone-contrast words of a syllable set by a given talker, they completed a corresponding 16-trial quiz. In each quiz trial, they heard one target word and had to choose the correct item in the display of four drawings. Correct answers were followed by a green tick on the screen and incorrect answers by a red cross mark. If the participant gave no response within 5.5 s, a reminder appeared asking them to respond more quickly. Following wrong or no answers, the correct drawing and audio word were presented to facilitate tone-word training (Laméris et al., 2024). Participants then were trained and quizzed on the 16 words produced by each of the other three talkers of the set, with the same cycle as shown in the second and third rectangles in the left panel of Figure 3.

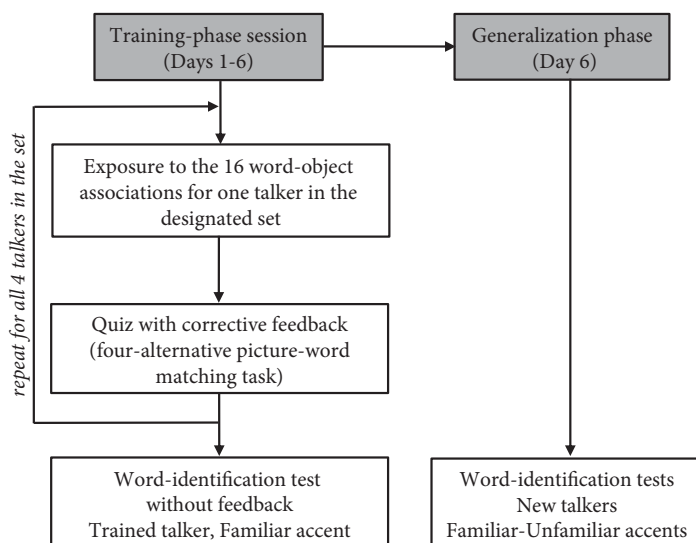


Figure 3. Summary of the experimental protocol.

To track learning progress across the six sessions, learners completed tone-word identification tests at the end of each session with 64 randomized trials (16 words $\times$ 4 tokens) (Figure 3, bottom left). The four tokens per word were one each from the four talkers of a given session to maintain talker variation, and token per talker per word was counterbalanced across learners to include token variation. On a given trial, learners saw the 16 pictures, which were displayed on a 4×4 answer matrix (as shown in Figure 2) on the participant's computer screen, where syllables and tones were blocked by row and column, respectively, and they heard a token depicting one of the 16 pictures. The item positions in the answer display were counterbalanced across participants but were constant across sessions for each participant.

After hearing the word, learners were instructed to identify the word by clicking on the correct picture. To ensure each tone-word was heard in full before an identification judgment was made, learners had to activate each trial by pressing and holding a mouse-click on a centered button on the screen until they heard the full stimulus. If they released it too soon, the trial ended automatically with a reminder to respond *after* hearing the word. Terminated trials (139 occurrences, or 0.7% of all trials across learners and sessions) were repeated in a random order at the end of each test. In contrast to the quizzes within training sessions, identification tests provided no feedback. To encourage careful thinking, there was no timeout, but learners responded within 5.5 s on average.

Immediately after the word-identification test of the final training session, learners completed generalization tests to new talkers of the familiar Beijing and the unfamiliar Shanghai accents, in that order, following the same word-identification procedure (Figure 3, bottom right). Only 38 trials (0.6%) were terminated by premature mouse click release. These two tests took about 7 min to complete.

Participants completed the training and testing online using their own laptop computer. The experimenter supervised the sessions via ZOOM to ensure data quality. The experiment was controlled using E-Prime Go (Psychology Software Tools, Inc.), which set the display resolution at 1920×1080 . Stimuli were presented via the

participant's wired earphones or headphones, which they adjusted to comfortable loudness during a pre-test practice session.

Statistical analysis

Bayesian mixed-effects models were used to analyze the perceptual data, because they account for cross-participant and cross-item effects, and provide measures of evidence regarding the hypotheses under investigation. This is particularly valuable in our study, where graded evidence for learning effects across sessions is more informative than simply determining whether effects are statistically significant. All models were implemented using the *brms* package in R (v. 4.4.0) (R Core Team, 2024).

We used a Bernoulli logit link to predict correct versus incorrect identification based on the training group, the session, and their interaction for learning and generalization data. This approach is well-suited for analyzing categorical response data and allows for interpretable estimates of changes in accuracy over time. We also included predictors for trial number to account for fatigue effects, and how often a specific stimulus had been presented to account for token familiarity (Herff *et al.*, 2018). All models included random intercepts for participant. Continuous variables were standardized to a mean of 0 and a standard deviation of 1. We used weakly informative priors in the form of t distributions with a mean of 0, a standard deviation of 1, and 3 degrees of freedom (Gelman *et al.*, 2008). R -hat was used to assess convergence, and all R -hats were 1 throughout (Bürkner, 2017). This approach has commonly been applied in auditory perception studies (Escudero *et al.*, 2023; Milne & Herff, 2020) both within and outside the linguistics domain.

To understand the separate contributions of tones and syllables in the current design, we scored tone and syllable accuracy separately: a response was counted as correct for the syllable even if the tone was incorrect, and vice versa. We then applied the same modeling architecture as above to the separated tone accuracy and syllable accuracy data, but added a *ToneOrSyllable* factor that interacted with session and accent group. We used the package *marginaleffects* to identify posterior predictions of the model and performed hypothesis tests on them to assess the factors of interest and their interactions. For these hypothesis tests, we reported regression coefficients (β), their estimate error (EE_β), evidence ratios (ERs) for directional effects ($\text{Odds}(\beta > 0 \text{ or } \beta < 0)$), and the corresponding posterior probabilities (Post.Prob). Following Milne and Herff (2020), ERs greater than 19 were taken to indicate strong evidence for a directed hypothesis. This threshold reflects the relative support for the hypothesized effect over its alternative, with larger values indicating stronger evidence. Full posterior summaries are provided in the Supplementary Materials. For convenience, we denoted effects with “*” that can be considered “significant” at an $\alpha = .05$ level.

Results

Mandarin minimal-tone-contrast word learning and identification

Overall word learning and identification

The Bayesian models of the identification test data from the six training sessions provided strong evidence that both the single- ($\beta = 0.41$, $EE_\beta = 0.01$, $\text{Odds}(\beta > 0) = 9999^*$, $\text{Post.Prob} = 1.00$) and multiple-accent ($\beta = 0.32$, $EE_\beta = 0.01$, $\text{Odds}(\beta > 0) > 9999^*$, $\text{Post.Prob} = 1.00$) groups improved from the first ($M_{\text{single-accent}} = 44.54\%$, $SD = 22.50$ vs. $M_{\text{multiple-accent}} = 55.33\%$, $SD = 23.33$) to the final ($M_{\text{single-accent}} = 85.75\%$, $SD = 13.52$ vs. $M_{\text{multiple-accent}} = 86.96\%$, $SD = 9.36$) training session (see Figure 4).

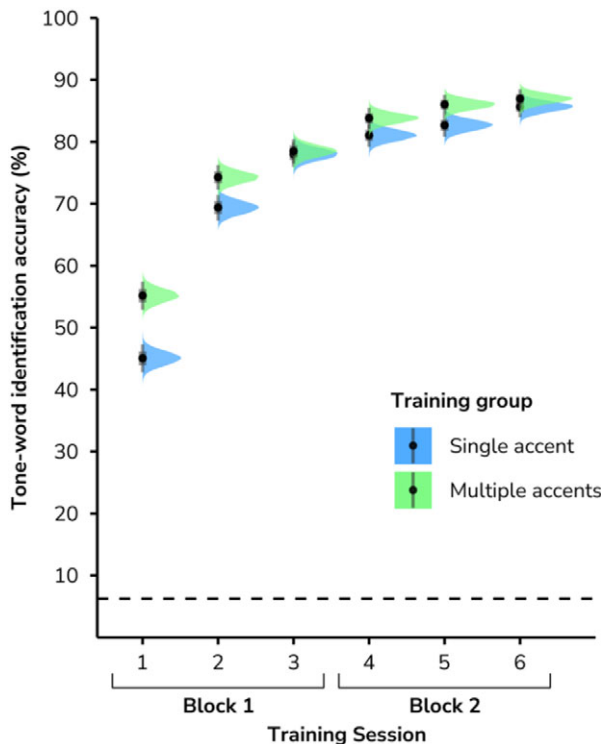


Figure 4. Posterior distributions of expected predictions for training effects on tone-word identification accuracy by single-accent versus multiple-accent groups. The dashed line indicates chance level.

There was also strong evidence illustrating that both groups improved across sessions in Block 1 (Sessions 1–3, $M_{\text{single-accent}} = 33.62\%$, $SD = 22.52$ vs. $M_{\text{multiple-accent}} = 23.04\%$, $SD = 18.88$; Single-accent: $\beta = 0.33$, $EE_{\beta} = 0.02$, $\text{Odds}(\beta > 0) > 9999^*$, $\text{Post.Prob} = 1.00$; Multiple-accent: $\beta = 0.23$, $EE_{\beta} = 0.02$, $\text{Odds}(\beta > 0) > 9999^*$, $\text{Post.Prob} = 1.00$), and across sessions in Block 2 (Sessions 4–6, $M_{\text{single-accent}} = 4.67\%$, $SD = 6.34$ vs. $M_{\text{multiple-accent}} = 3.21\%$, $SD = 15.62$; Single-accent: $\beta = 0.05$, $EE_{\beta} = 0.01$, $\text{Odds}(\beta > 0) > 9999^*$, $\text{Post.Prob} = 1.00$; Multiple-accent: $\beta = 0.03$, $EE_{\beta} = 0.01$, $\text{Odds}(\beta > 0) = 203.08^*$, $\text{Post.Prob} = 1.00$). The Bayesian models also provided strong evidence that, except for Sessions 3 and 6 (the final sessions of Blocks 1 and 2), the multiple-accent group performed better than the single-accent group (all ERs > 19 for Sessions 1, 2, 4, and 5), indicating a consistent performance advantage for the multiple-accent group across training sessions. Both groups had reached asymptotic performance in each training block by the second session of that block, with small additional improvement in the final session of each block.

Performance split by tone and syllable

To examine whether tone and/or syllable learning contributed to group differences, the responses were rescored to give separate tone and syllable identification accuracies. These overall tone and syllable scores were then broken down by individual tones and syllables to assess the effects of accent variability on specific tone and syllable learning.

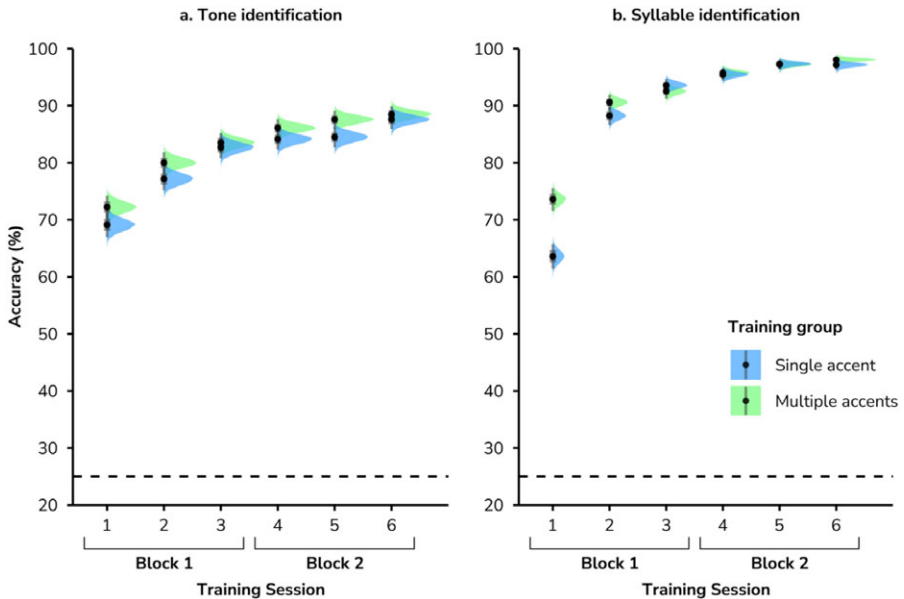


Figure 5. Posterior distributions for training effects on tone and syllable accuracy by single-accent versus multiple-accent groups. The dashed line indicates chance level.

As shown in Figure 5, both groups improved tone identification gradually from the first session ($M_{\text{single-accent}} = 69.71\%$, $SD = 15.07$ vs. $M_{\text{multiple-accent}} = 71.96\%$, $SD = 18.05$) to the sixth session ($M_{\text{single-accent}} = 87.50\%$, $SD = 11.96$ vs. $M_{\text{multiple-accent}} = 88.46\%$, $SD = 8.94$) for approximately 17 percentage points ($M_{\text{single-accent}} = 17.79\%$, $SD = 14.07$; $M_{\text{multiple-accent}} = 16.50\%$, $SD = 15.40$), with greater improvement in Block 1 ($M_{\text{single-accent}} = 12.92\%$, $SD = 13.22$; $M_{\text{multiple-accent}} = 11.58\%$, $SD = 12.87$) than Block 2 ($M_{\text{single-accent}} = 3.50\%$, $SD = 6.40$; $M_{\text{multiple-accent}} = 2.33\%$, $SD = 11.26$). Again, we observed strong evidence for these improvements (all ERs > 19). We also observed strong evidence, again, that the multiple-accent group performed better than the single-accent group in all sessions except for the final session in each block (Sessions 3 and 6).

Similar to tone identification, we observed strong evidence that both groups improved syllable identification from Sessions 1 to 6 as well as within block improvements (all ERs > 19). There was also strong evidence that the multiple-accent group performed better initially, in Session 1 ($M_{\text{single-accent}} = 62.50\%$, $SD = 23.04$ vs. $M_{\text{multiple-accent}} = 73.88\%$, $SD = 23.00$; $\beta = 0.10$, $EE_{\beta} = 0.01$, Odds($\beta > 0$) > 9999*, Post.Prob = 1.00). The syllable analysis broken down by individual syllables (see Figure 6) showed that the multiple-accent group outperformed the single-accent group in identifying all four syllables (all ERs > 19). Both groups achieved over 88% mean syllable accuracy by Session 2, i.e., their performance asymptoted well before the final session even though it continued to show small improvements from Session 2 to Session 6 ($M_{\text{single-accent}} = 97.17\%$, $SD = 5.55$ vs. $M_{\text{multiple-accent}} = 97.96\%$, $SD = 2.61$). Together, these findings suggest that improvements in tone-word learning across training sessions were driven primarily by gains in tone identification rather than in syllable learning.

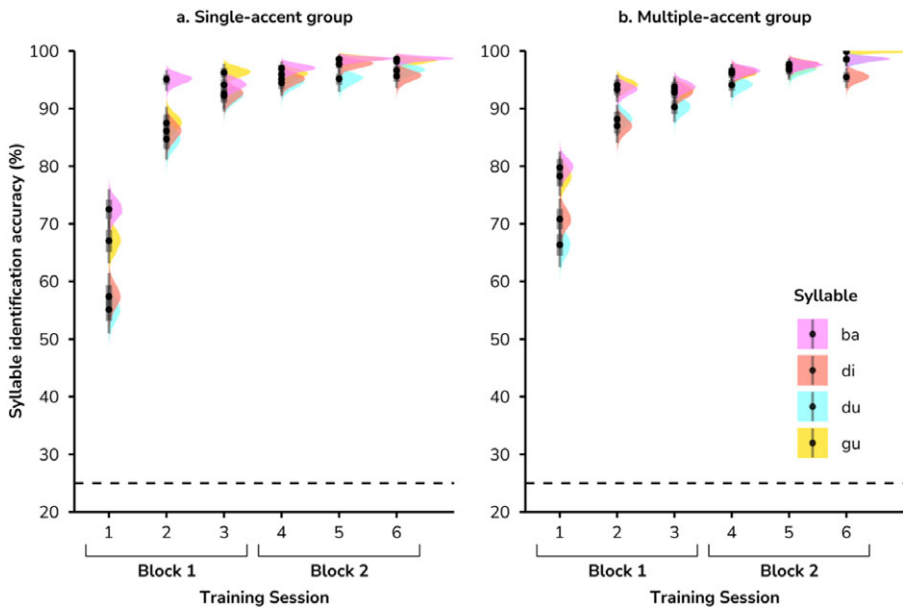


Figure 6. Posterior distributions for training effects on syllable identification accuracy by syllable and accent group. The dashed line indicates chance level.

For identification accuracy of the four tones (see Figure 7), both groups showed improvements from the first to the sixth training session for each tone (all ERs > 19). While the single-accent group outperformed the multiple-accent group in identifying dipping tone after Session 1 ($M_{\text{single-accent}} = 84.38\%$, $SD = 36.36$ vs. $M_{\text{multiple-accent}} = 78.13\%$, $SD = 41.39$; $\beta = 0.06$, $EE_{\beta} = 0.02$, Odds($\beta > 0$) = 231.56*, Post.Prob = 1.00), the multiple-accent group outperformed the single-accent group in identifying rising ($M_{\text{single-accent}} = 61.72\%$, $SD = 48.67$ vs. $M_{\text{multiple-accent}} = 68.23\%$, $SD = 46.62$; $\beta = 0.06$, $EE_{\beta} = 0.03$, Odds($\beta > 0$) = 49.76*, Post.Prob = 0.98) and falling ($M_{\text{single-accent}} = 65.36\%$, $SD = 47.64$ vs. $M_{\text{multiple-accent}} = 71.61\%$, $SD = 45.15$; $\beta = 0.06$, $EE_{\beta} = 0.03$, Odds($\beta > 0$) = 43.44*, Post.Prob = 0.98) tones. The multiple-accent group retained higher rising-tone identification accuracy than the single-accent group until Session 5 ($M_{\text{single-accent}} = 77.86\%$, $SD = 41.57$ vs. $M_{\text{multiple-accent}} = 86.72\%$, $SD = 33.98$; $\beta = 0.09$, $EE_{\beta} = 0.03$, Odds($\beta > 0$) = 2499.00*, Post.Prob = 1.00). However, by Session 6, both groups reached equivalent performance on the four tones. Across the four tones, we observed strong evidence that both groups demonstrated the highest identification accuracy for the dipping tone after Session 1, and higher accuracy for the dipping tone than for the level tone at Session 6 (all ERs > 19). These results suggest that dipping tone is learned more easily than the other three tones when paired with lexical meanings.

Generalization to new talkers of a familiar and an unfamiliar accent

We focused here on the split accuracy for tones and for syllables (see Figure 8), as we were interested in generalization of phonological abstraction for lexical tones and syllables to new talkers from Beijing and Shanghai. As a reminder, both the single-accent group and the multiple-accent group had been trained on the Beijing accent, and

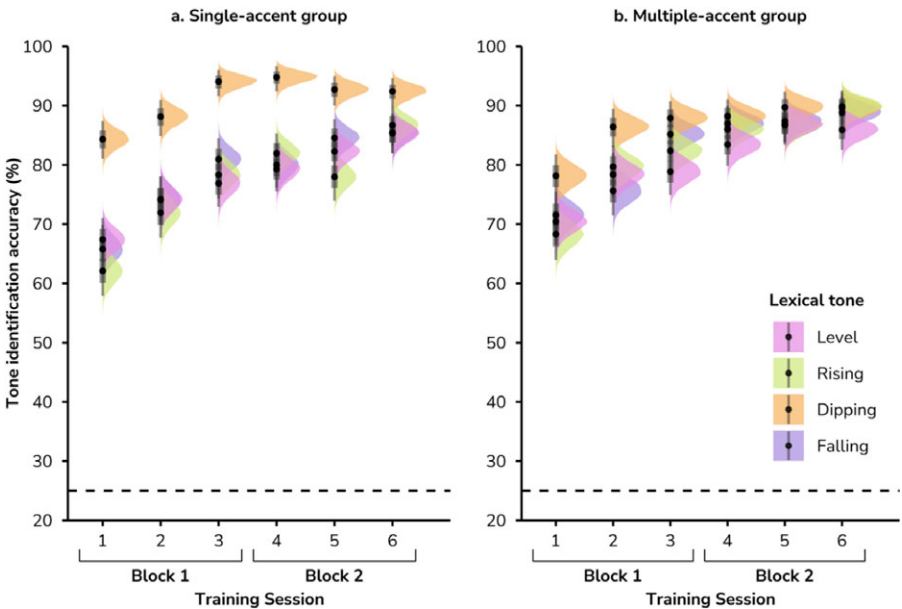


Figure 7. Posterior distributions for training effects on tone identification by tone and accent group. The dashed line indicates chance level.

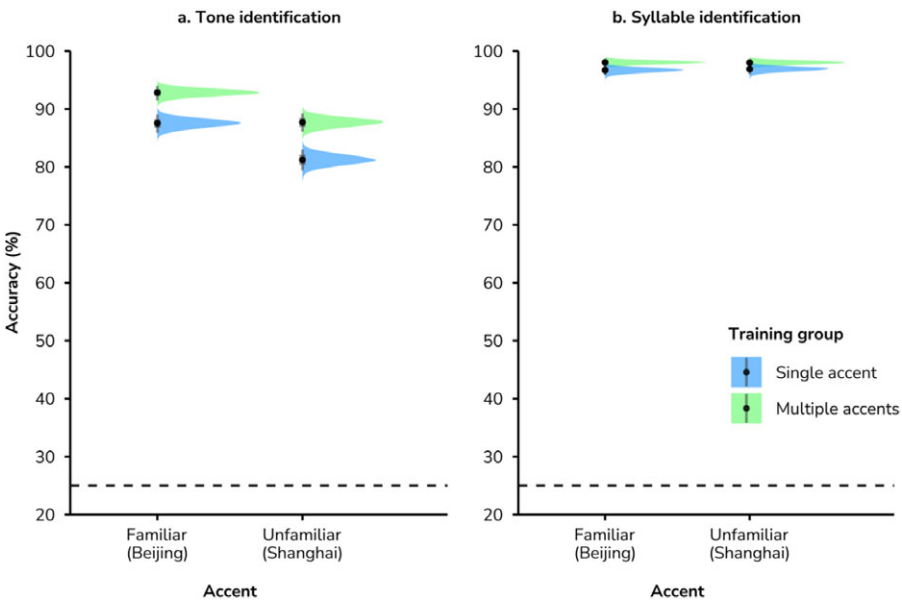


Figure 8. Posterior distributions for tone- and syllable-identification accuracy in generalization tests to new talkers with a familiar Beijing versus an unfamiliar Shanghai accent. The dashed line indicates chance level.

neither group was trained on the Shanghai accent. The data analysis method was similar to that in Mandarin tone-word identification tests.

There was strong evidence that exposure to multiple accents during word training resulted in better tone generalization than single-accent training for both the new talker with the familiar Beijing accent (see Figure 8a; $M_{\text{single-accent}} = 87.46\%$, $SD = 11.36$ vs. $M_{\text{multiple-accent}} = 92.88\%$, $SD = 5.80$; $\beta = 0.05$, $EE_{\beta} = 0.01$, $\text{Odds}(\beta > 0) > 9999^*$, $\text{Post. Prob} = 1.00$), and the new talker with the unfamiliar Shanghai accent ($M_{\text{single-accent}} = 81.08\%$, $SD = 13.35$ vs. $M_{\text{multiple-accent}} = 87.67\%$, $SD = 9.43$; $\beta = 0.06$, $EE_{\beta} = 0.01$, $\text{Odds}(\beta > 0) > 9999^*$, $\text{Post. Prob} = 1.00$). There was also strong evidence that both groups showed slightly better generalization to the new talker with the familiar Beijing accent than to the new talker with the unfamiliar Shanghai accent (Single-accent: $\beta = 0.06$, $EE_{\beta} = 0.01$, $\text{Odds}(\beta > 0) = 9999^*$, $\text{Post. Prob} = 1.00$; Multiple-accent: $\beta = 0.05$, $EE_{\beta} = 0.01$, $\text{Odds}(\beta > 0) = 9999^*$, $\text{Post. Prob} = 1.00$).

For syllable processing, both groups achieved 95% accuracy (see Figure 8b) in generalizations to new talkers of the familiar Beijing accent ($M_{\text{single-accent}} = 96.54\%$, $SD = 5.47$ vs. $M_{\text{multiple-accent}} = 97.96\%$, $SD = 3.62$) and of the unfamiliar Shanghai accent ($M_{\text{single-accent}} = 96.83\%$, $SD = 6.65$ vs. $M_{\text{multiple-accent}} = 97.88\%$, $SD = 2.91$). Accordingly, we did not find significant evidence that participants performed better in generalization of syllables to the new talker of the familiar accent than to the new talker of the unfamiliar accent. Nevertheless, similar to tone identification, there was strong evidence that the multiple accent group performed better than the single accent group in the two generalization tests (Beijing: $\beta = 0.01$, $EE_{\beta} = 0.01$, $\text{Odds}(\beta > 0) = 242.90^*$, $\text{Post. Prob} = 1.00$; Shanghai: $\beta = 0.01$, $EE_{\beta} = 0.01$, $\text{Odds}(\beta > 0) = 52.48^*$, $\text{Post. Prob} = 0.98$).

Figure 9 shows group differences broken down by specific tone. For generalization to the new talker with a familiar accent, the multiple-accent group outperformed the

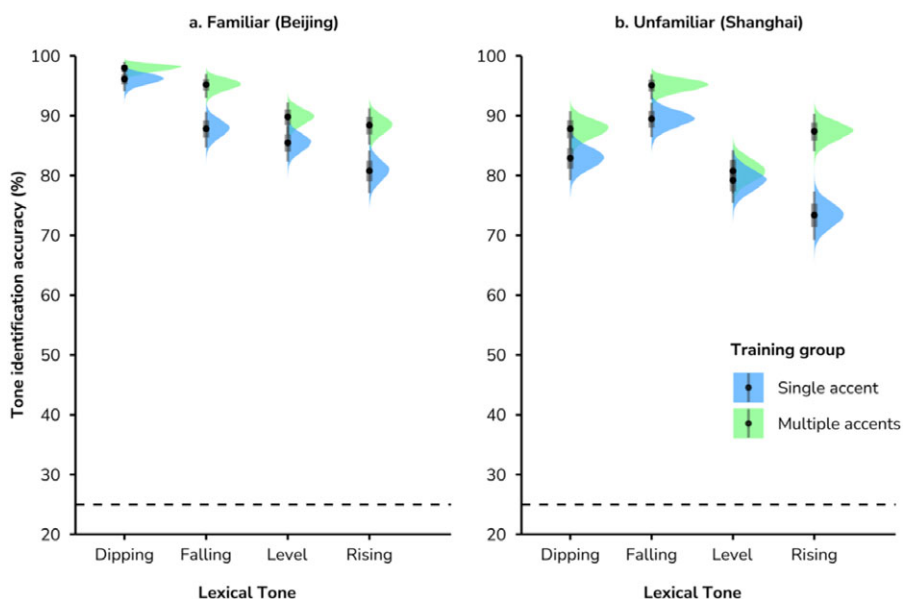


Figure 9. Posterior distributions for tone-identification accuracy in generalization tests involving new talkers with familiar Beijing versus unfamiliar Shanghai accents, split by tone. The dashed line indicates chance level.

single-accent group in identifying all four tones, i.e., level ($M_{\text{single-accent}} = 85.16\%$, $SD = 35.60$ vs. $M_{\text{multiple-accent}} = 89.84\%$, $SD = 30.25$), rising ($M_{\text{single-accent}} = 80.73\%$, $SD = 39.49$ vs. $M_{\text{multiple-accent}} = 88.28\%$, $SD = 32.21$), dipping ($M_{\text{single-accent}} = 96.61\%$, $SD = 18.11$ vs. $M_{\text{multiple-accent}} = 97.92\%$, $SD = 14.30$), and falling ($M_{\text{single-accent}} = 87.50\%$, $SD = 33.12$ vs. $M_{\text{multiple-accent}} = 95.31\%$, $SD = 21.16$) tones. For generalization to the new talker with an unfamiliar accent, the multiple accent group showed higher identification accuracy for the rising ($M_{\text{single-accent}} = 73.18\%$, $SD = 44.36$ vs. $M_{\text{multiple-accent}} = 87.50\%$, $SD = 33.12$), dipping ($M_{\text{single}} = 82.55\%$, $SD = 38.00$ vs. $M_{\text{multiple}} = 87.76\%$, $SD = 32.82$), and falling ($M_{\text{single-accent}} = 89.58\%$, $SD = 30.59$ vs. $M_{\text{multiple-accent}} = 95.05\%$, $SD = 21.71$) tones. All effects were strong (all ERs > 19). We also found that both groups demonstrated higher identification accuracy for the level and falling tones when generalizing to the new talker with the familiar accent than to the one with an unfamiliar accent (all ERs > 19). In addition, the single-accent group showed better identification of the rising tone for generalization to the familiar-accent talker than to the unfamiliar-accent talker ($\beta = 0.07$, $EE_{\beta} = 0.03$, Odds($\beta > 0$) = 255.41*, Post.Prob = 1.00). Across the four tones, both groups achieved the highest identification accuracy for the dipping tone when generalizing to the new talker with a familiar accent, whereas the highest accuracy for generalization to the new talker with an unfamiliar accent was observed for the falling tone. All effects showed strong evidence (all ERs > 19).

For syllable contributions to group differences (see Figure 10), there was strong evidence that in generalization to the new talker of the familiar Beijing accent the multiple-accent group performed better than the single-accent group in identifying words with the syllables /du/ ($M_{\text{single-accent}} = 95.31\%$, $SD = 21.16$ vs. $M_{\text{multiple-accent}} = 98.44\%$, $SD = 12.42$; $\beta = 0.03$, $EE_{\beta} = 0.01$, Odds($\beta > 0$) = 195.08*, Post.Prob = 0.99) and /gu/ ($M_{\text{single-accent}} = 96.61\%$, $SD = 18.11$ vs. $M_{\text{multiple-accent}} = 98.44\%$, $SD = 12.42$; $\beta = 0.01$,

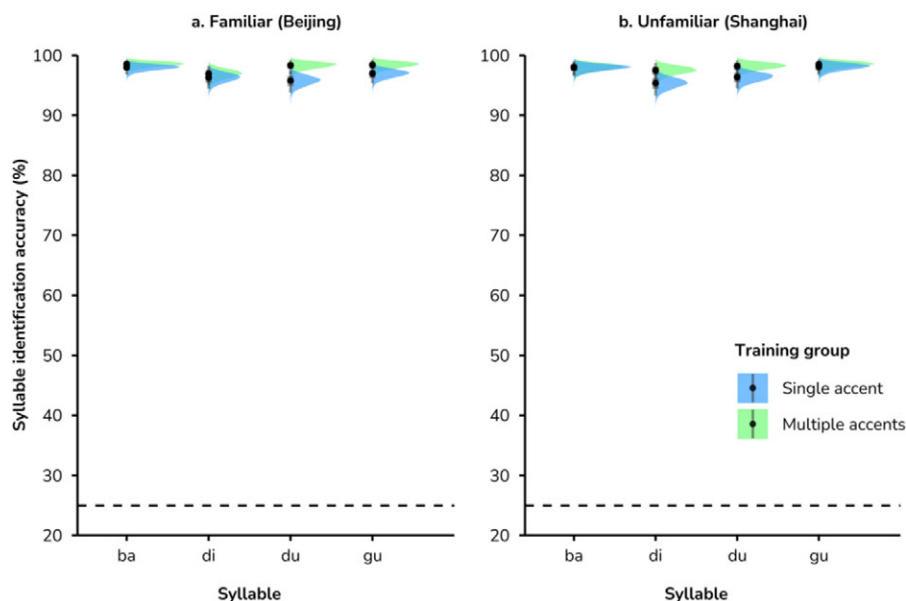


Figure 10. Posterior distributions for syllable-identification accuracy in generalization tests involving new talkers with familiar Beijing versus unfamiliar Shanghai accents, split by syllable. The dashed line indicates chance level.

$EE_{\beta} = 0.01$, Odds($\beta > 0$) = 20.46*, Post.Prob = 0.95). In generalization to the new talker of the unfamiliar Shanghai accent, the multiple-accent group outperformed the single-accent group in identifying words with the syllables /du/ ($M_{\text{single-accent}} = 96.35\%$, $SD = 18.77$ vs. $M_{\text{multiple-accent}} = 98.18\%$, $SD = 13.40$; $\beta = 0.02$, $EE_{\beta} = 0.01$, Odds($\beta > 0$) = 26.55*, Post.Prob = 0.96) and /di/ ($M_{\text{single-accent}} = 95.05\%$, $SD = 21.71$ vs. $M_{\text{multiple-accent}} = 97.66\%$, $SD = 15.15$; $\beta = 0.02$, $EE_{\beta} = 0.01$, Odds($\beta > 0$) = 26.62*, Post.Prob = 0.96).

Discussion

The present study investigated the effects of accent variability on monolingual English perceivers' acquisition of Mandarin minimal-tone-contrast words. Participants completed six sessions of tone-word training, and their learning trajectories were tracked using word-identification tests administered after each session. They subsequently completed two generalization tests to new talkers with a familiar Beijing and an unfamiliar Shanghai accent.

Both the single-accent and multiple-accent groups reached around 85% accuracy by the final training session having asymptoted by the third session, despite completing fewer training sessions (six in our study) than the nine sessions used in Chandrasekaran et al. (2010) and Wong and Perrachione (2007). These results indicate that our modified training program was both effective and efficient for tone-word training, further supported by the high levels of dipping-based word identification observed in both groups. Notably, the multiple-accent group outperformed the single-accent group during the first two sessions of each training block. This early advantage was primarily driven by tone identification, as both groups reached 88+% accuracy in syllable identification by the second session. Because the multiple-accent group had already been exposed to accent variability during the initial two sessions of Block 1, their superior tone identification in those sessions (as well as the first two sessions of Block 2) suggests that exposure to multiple accents facilitated the extraction of more stable abstract representations of lexical tones. In other words, accent variability appears to promote the emergence of phonological constancy at an early stage of tone-word learning. However, this advantage diminished by the third session, as both groups reached nearly 80% accuracy regardless of accent condition. This convergence suggests that talker variability alone, experienced by both groups, was highly effective and that additional accent variability provided an early but relatively modest enhancement. Our prediction regarding the effect of accent variability on tone-word learning was therefore modestly supported. The high identification accuracy achieved by the final session of Block 2 in training further suggested that both groups had established abstract phonological representations of the lexical tones, which were subsequently evaluated in the two generalization tests.

Both groups maintained ceiling-level syllable identification in the two generalization tests, suggesting that the tone-naïve English learners had established robust abstract phonological representations of the four trained Mandarin syllables. Thus, overall word identification accuracy primarily depended on tone identification, consistent with previous studies on Mandarin tone-word learning by English learners (Hao, 2024; Pelzl et al., 2021). As predicted, the multiple-accent group outperformed the single-accent group in generalizing to the new talkers with both a familiar Beijing and an unfamiliar Shanghai accent. This advantage indicates that exposure to accent variability during tone-word training facilitated a more robust emergence of phonological constancy across talkers and accents. This pattern was further supported by tone-specific analyses: the multiple-accent group showed superior identification of all four tones in

the familiar-accent condition and of three tones (all except the level tone) in the unfamiliar-accent condition. Interestingly, across the two generalization tests, both groups showed an overall advantage in the familiar-accent condition over the unfamiliar-accent condition, suggesting that phonological constancy had emerged but remained sensitive to accent familiarity. This finding is consistent with accent adaptation studies (Bradlow & Bent, 2008; Clarke & Garrett, 2004), as well as with meta-analytic evidence on the generalization effects of high-variability phonetic training (Uchihara *et al.*, 2025). Although learners were able to generalize to new talkers, performance was facilitated when the accent matched previously experienced acoustic distributions. This pattern indicates that tone representations were abstract enough to support cross-talker generalization, yet not fully accent-invariant, particularly in the single-accent group, reflecting a gradual emergence of phonological constancy.

The present study advanced the perceptual attunement account (Best, 2015) by demonstrating that structured indexical variability plays a critical role in L2 tone-word learning. Because F0 movement in lexical tones is intrinsically tied to talker and accent variation (Li *et al.*, 2020a, 2020b), successful tone acquisition requires learners to abstract invariant cues from notably variable input. Our findings showed that English learners achieved such abstraction across both talker and accent differences, extending the scope of perceptual attunement beyond consonant and vowel perception (Best *et al.*, 2009) to lexical tone processing. Accent variability conferred greater generalization benefits than talker variability alone, underscoring the privileged role of higher-level indexical variation (Tyler & Best, *in preparation*; see also Foulkes & Hay, 2015) in promoting phonological constancy. The observed generalization patterns provided strong support for hybrid models of word recognition (Cutler & Weber, 2007; Goldinger, 2007), indicating that L2 tone representations simultaneously retain indexical detail and undergo progressive abstraction. These results indicate that phonological constancy in L2 tone learning is not the product of mere exposure but emerges through handling structured variability in the input as a fundamentally graded and dynamic process.

Beyond the establishment of new tone categories, the extent to which these representations generalize may depend on the acoustic relationship between training and test accents. Previous work demonstrates that cross-talker generalization is modulated by acoustic similarity among talker voices, with greater similarity facilitating transfer (Xie *et al.*, 2021; Xie & Myers, 2017). Extending this logic to the accent level, if the Shanghai accent occupies an intermediate acoustic position relative to the training accents (see Figure 1), generalization performance may reflect graded similarity rather than categorical transfer. From this perspective, the observed multiple-accent advantage may represent a conservative estimate of variability effects: larger acoustic distances between training and test accents could further amplify the role of structured variability in promoting phonological constancy. Future research that systematically manipulates accent distance (see Bent *et al.*, 2024) in lexical tones would clarify how similarity at different levels of the indexical hierarchy (Tyler & Best, *in preparation*; see also Foulkes & Hay, 2015) constrains variability-driven abstraction. Importantly, taken together, the generalization findings indicate the emergence of phonological constancy in L2 tone-word learning, suggesting that learners established abstract tone representations at the sub-lexical level. While the present study demonstrated the emergence of such abstraction across talkers and accents, future research could more directly assess this change by examining tone identification and discrimination of untrained tone-words following tone-word training (Chandrasekaran *et al.*, 2010; see also Li *et al.*, 2022, 2023a).

Meanwhile, long-term stability of phonological abstraction for tones in lexical contexts remains an open question. While a recent study has provided meta-analytic evidence on the long-term retention effects of high-variability phonetic training (Uchihara et al., 2025), delayed post-tests would offer a complementary approach to examining the durability of accent-variability-driven abstraction over time. Investigating how phonological constancy stabilizes or evolves beyond the initial learning phase would further refine theoretical models of L2 lexical tone acquisition.

The findings in this study contribute to our understanding of variability-based learning in L2 tone perception, and several important questions remain for future research. The present study used isolated monosyllabic minimal-tone-contrast words to ensure experimental control. Mandarin tones in natural speech interact with segmental context, prosodic structure, coarticulation, and intonation, and processes such as tone sandhi (see Duanmu, 2007, for a review), all of which may alter their acoustic realization and perceptual salience. It remains to be determined whether the benefits of high-accent variability extend to sentence- or discourse-level contexts. In addition, because this study focused exclusively on monolingual English listeners, whose L1 lacks lexical tone, it remains unclear whether the facilitative effects of accent variability are specific to non-tone language learners or whether they generalize across learners with different L1 backgrounds. Cross-linguistic comparisons among tone-language learners (e.g., Cantonese, Thai), pitch-accent-language learners (e.g., Swedish, Japanese), and non-tone-language learners (Laméris et al., 2024) would clarify whether prior experience with lexical pitch distinctions modulates sensitivity to accent variability in lexical tones and the benefits derived from high-accent-variability input.

In sum, the present findings demonstrated that structured accent variability systematically enhances the robustness of L2 tone representations at the sub-lexical level. Rather than yielding fully invariant categories, accent variability promoted progressively abstract yet indexically informed representations. These results provided direct evidence that perceptual attunement in adult L2 tone learning is variability-driven, gradient, and fundamentally shaped by the structure of the input.

Data availability statement. The experiment in this article earned Open Data and Pre-Registered badges for transparent practices. The data are available at <https://doi.org/10.17605/OSF.IO/C6SBM>.

Acknowledgments. The authors would like to gratefully acknowledge insightful comments from three anonymous reviewers, as well as Handling Editor Kazuya Saito and Editor Luke Plonsky, on earlier versions of the manuscript. This study was supported by a joint PhD scholarship from Western Sydney University (WSU) and the China Scholarship Council (CSC), as well as by Research Training Funds from the MARCS Institute for Brain, Behaviour and Development at WSU. It was also partly funded by two Australian Linguistic Society Research Grants awarded to the first author. Co-author Herff's work on this report was supported by an Australian Research Council (ARC) Discovery Early Career Research Award (DECRA, DE220100961) and a Sydney Horizon Fellowship (University of Sydney). The authors would like to thank CareerHub at Australian Catholic University, Australian National University, The University of Melbourne, The University of Western Australia, and Western Sydney University for assisting with online recruitment of participants during the Covid-19 pandemic. The authors would also like to thank Johnson Chen, Dr. Chris Wang, and Dr. Gabriele Cecchetti for their technical support at the MARCS Institute for Brain, Behaviour and Development, as well as Professor Jinsong Zhang at the Speech Acquisition and Intelligent Technology (SAIT) Lab, Beijing Language and Culture University, China, for granting us access to the lab's professional recording booth. Above all, we are grateful to the participants for their time and involvement. Parts of these findings were presented at the 20th International Congress of the Phonetic Science (ICPhS), Prague, Czech Republic, in August 2023 (Li et al., 2023b).

Competing interests. The authors declare none.

References

- Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition*, 27(3), 387–414. <https://doi.org/10.1017/S0272263105050175>
- Barcroft, J., & Sommers, M. S. (2014). Effects of variability in fundamental frequency on L2 vocabulary learning: A comparison between learners who do and do not speak a tone language. *Studies in Second Language Acquisition*, 36(3), 423–449. <https://doi.org/10.1017/S0272263113000582>
- Bent, T., Henry, M., Holt, R., & Lind-Combs, H. (2024). Relating pronunciation distance metrics to intelligibility across English accents. *Journal of Phonetics*, 107. <https://doi.org/10.1016/j.wocn.2024.101357>
- Best, C. T. (2015). Devil or angel in the details?: Perceiving phonetic variation as information about phonological structure. In J. Romero & M. Riera (Eds.), *Current issues in linguistic theory* (Vol. 335, pp. 3–32). John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.335.01bes>
- Best, C. T. (2019). The diversity of tone languages and the roles of pitch variation in non-tone languages: Considerations for tone perception research. *Frontiers in Psychology*, 10, 364. <https://doi.org/10.3389/fpsyg.2019.00364>
- Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, 20(5), 539–542. <https://doi.org/10.1111/j.1467-9280.2009.02327.x>
- Boersma, P., & Weenink, D. (2018). *Praat: Doing Phonetics by Computer* (Version 6.0.43) [Computer software]. <https://www.fon.hum.uva.nl/praat/>
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729. <https://doi.org/10.1016/j.cognition.2007.04.005>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Burnham, D., Brooker, R., & Reid, A. (2015). The effects of absolute pitch ability and musical training on lexical tone perception. *Psychology of Music*, 43(6), 881–897. <https://doi.org/10.1177/0305735614546359>
- Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, 128(1), 456–465. <https://doi.org/10.1121/1.3445785>
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116(6), 3647–3658. <https://doi.org/10.1121/1.1815131>
- Cutler, A., & Weber, A. (2007). Listening experience and phonetic-to-lexical mapping in L2. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 43–48).
- Dong, H., Clayards, M., Brown, H., & Wonnacott, E. (2019). The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones. *PeerJ*, 7, e7191. <https://doi.org/10.7717/peerj.7191>
- Duanmu, S. (2007). *The phonology of Standard Chinese* (2nd ed). Oxford University Press.
- Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Platsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, 71(4), 808–816. <https://doi.org/10.1080/17470218.2017.1310261>
- Escudero, P., Smit, E. A., & Angwin, A. J. (2023). Investigating orthographic versus auditory cross-situational word learning with online and laboratory-based testing. *Language Learning*, 73(2), 543–577. <https://doi.org/10.1111/lang.12550>
- Fichtner, F., Barcroft, J., Sommers, M., & Olejarczuk, P. (2025). Effects of sociophonetic variability on L2 vocabulary learning. *Studies in Second Language Acquisition*, 47(4), 1071–1095. <https://doi.org/10.1017/S0272263125100648>
- Foulkes, P., & Hay, J. (2015). The emergence of sociophonetic structure. In W. O'Grady & B. MacWhinney (Eds.), *The handbook of language emergence* (pp. 292–313). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118346136.ch13>
- Fromkin, V. (Ed.). (1978). *Tone: A linguistic survey*. Academic Press.
- Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11(2), 149–175. [https://doi.org/10.1016/S0095-4470\(19\)30813-7](https://doi.org/10.1016/S0095-4470(19)30813-7)
- Ge, Y., Correia, S., Fernandes, J., Hanson, K., Rato, A., & Rebuschat, P. (2026). Does phonetic training benefit word learning? *Studies in Second Language Acquisition*, 1–28. <https://doi.org/10.1017/s0272263125101459.sm001>

- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4), 1360–1383. <https://doi.org/10.1214/08-AOAS191>
- Goldinger, S. (2007). A complementary-systems approach to abstract and episodic speech perception. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 49–53).
- Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40(2), 269–279. <https://doi.org/10.1016/j.wocn.2011.11.001>
- Hao, Y.-C. (2018). Second language perception of Mandarin vowels and tones. *Language and Speech*, 61(1), 135–152. <https://doi.org/10.1177/0023830917717759>
- Hao, Y.-C. (2024). Phonolexical processing of Mandarin segments and tones by English speakers at different Mandarin proficiency levels. *Second Language Research*, 40(3), 533–557. <https://doi.org/10.1177/02676583231167790>
- Herff, S. A., Olsen, K. N., & Dean, R. T. (2018). Resilient memory for melodies: The number of intervening melodies does not influence novel melody recognition. *Quarterly Journal of Experimental Psychology* (2006), 71(5), 1150–1171. <https://doi.org/10.1080/17470218.2017.1318932>
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In *talker variability in speech processing* (1st ed., pp. 145–165). Academic Press. <https://doi.org/10.1002/9780470757024.ch15>
- Lam  ris, T. J., Llompart, M., & Post, B. (2024). Non-native tone categorization and word learning across a spectrum of L1 tonal statuses. *Bilingualism: Language and Cognition*, 27(4), 729–743. <https://doi.org/10.1017/S1366728923000871>
- Li, Y., Best, C. T., Tyler, M. D., & Burnham, D. (2020a). Regionally accented Mandarin lexical tones. *The Journal of the Acoustical Society of America*, 148(4_Supplement), 2474–2475. <https://doi.org/10.1121/1.5146856>
- Li, Y., Best, C. T., Tyler, M. D., & Burnham, D. K. (2020b). Tone variations in regionally accented Mandarin. In H. Meng, B. Xu, & T. F. Zheng (Eds.), *Proceedings of the 21st Annual Conference of the International Speech Communication Association* (pp. 4158–4162). <https://doi.org/10.21437/interspeech.2020-1235>
- Li, Y., Best, C. T., Tyler, M. D., & Burnham, D. K. (2022). L2-Mandarin regional accent variability during lexical tone word training facilitates naive English listeners' tone categorization and discrimination. In R. Billington (Ed.), *Proceedings of the 18th Australasian International Conference on Speech Science and Technology* (pp. 156–160).
- Li, Y., Tyler, M., Burnham, D., & Best, C. (2023a). L2-Mandarin regional accent variability during Mandarin tone-word training facilitates English listeners' subsequent tone categorizations. *Proceedings of the 24th Annual Conference of the International Speech Communication Association*, 4259–4263. <https://doi.org/10.21437/Interspeech.2023-145>
- Li, Y., Tyler, M., Burnham, D., & Best, C. (2023b). L2-Mandarin regional accent variability facilitates Mandarin-na  ve English listeners' learning of Mandarin tone-words. In R. Skarnitzl & J. Vol  n (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 2457–2461). GUARANT International.
- Liu, L., G  tz, A., Lorette, P., & Tyler, M. D. (2022). How tone, intonation and emotion shape the development of infants' fundamental frequency perception. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.906848>
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6), 1113–1126. https://doi.org/10.1207/s15516709cog0000_79
- Milne, A. J., & Herff, S. A. (2020). The perceptual relevance of balance, evenness, and entropy in musical rhythms. *Cognition*, 203, 104233. <https://doi.org/10.1016/j.cognition.2020.104233>
- Nespor, M., & Vogel, I. (2007). *Prosodic phonology: With a new foreword*. De Gruyter Mouton. <https://doi.org/10.1515/9783110977790>
- Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2021). Even in the best-case scenario L2 learners have persistent difficulty perceiving and utilizing tones in Mandarin: Findings from behavioral and event-related potentials experiments. *Studies in Second Language Acquisition*, 43(2), 268–296. <https://doi.org/10.1017/S027226312000039X>

- 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*, 461–472. <https://doi.org/10.1121/1.3593366>
- R Core Team. (2024). *R: The R Project for Statistical Computing* (Version 4.4.0) [Computer software]. <https://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. <https://doi.org/10.3389/fpsyg.2014.01318>
- 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(2), 195–221. <https://doi.org/10.1017/S0272263114000047>
- Tyler, M. D. (2021). Perceived phonological overlap in second-language categories: The acquisition of English /r/ and /l/ by Japanese native listeners. *Languages*, 6(1), 4. <https://doi.org/10.3390/languages6010004>
- Tyler, M. D., & Best, C. T. (in preparation). Perceptual Assimilation as microscope: How experience helps resolve both phonological contrasts and phonetic details. *Journal of Phonetics*.
- Uchihara, T., Karas, M., & Thomson, R. I. (2025). High variability phonetic training (HVPT): A meta-analysis of L2 perceptual training studies. *Studies in Second Language Acquisition*, 47(3), 794–827. <https://doi.org/10.1017/S0272263125100879>
- Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022). The effects of talker variability and frequency of exposure on the acquisition of spoken word knowledge. *Studies in Second Language Acquisition*, 44(2), 357–380. <https://doi.org/10.1017/S0272263121000218>
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28(4), 565–585. <https://doi.org/10.1017/S0142716407070312>
- Xie, X., Liu, L., & Jaeger, T. F. (2021). Cross-talker generalization in the perception of nonnative speech: A large-scale replication. *Journal of Experimental Psychology: General*, 150(11), e22–e56. <https://doi.org/10.1037/xge0001039>
- Xie, X., & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language*, 97, 30–46. <https://doi.org/10.1016/j.jml.2017.07.005>

Cite this article: Li, Y., Herff, S. A., Tyler, M. D., Burnham, D., & Best, C. T. (2026). Effects of high accent variability on monolingual English perceivers' acquisition of Mandarin minimal-tone-contrast words. *Studies in Second Language Acquisition*, 1–24. <https://doi.org/10.1017/S0272263126101740>