frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 24 July 2024
DOI 10.3389/feduc.2024.1392022

OPEN ACCESS

EDITED BY
Xin Wang,
Macquarie University, Australia

REVIEWED BY
Fei Chen,
Hunan University, China
Si Chen,
Hong Kong Polytechnic University, Hong Kong SAR, China

*CORRESPONDENCE
Jules Vonessen
✉ jsvonessen@ucdavis.edu

RECEIVED 26 February 2024
ACCEPTED 08 July 2024
PUBLISHED 24 July 2024

CITATION
Vonessen J and Zellou G (2024) Perception of Mandarin tones across different phonological contexts by native and tone-naive listeners. *Front. Educ.* 9:1392022.
doi: 10.3389/feduc.2024.1392022

Perception of Mandarin tones across different phonological contexts by native and tone-naive listeners

Jules Vonessen* and Georgia Zellou

Phonetics Laboratory, Department of Linguistics, University of California, Davis, Davis, CA, United States

Coarticulation is a type of speech variation where sounds take on phonetic properties of adjacent sounds. Listeners generally display perceptual compensation, attributing coarticulatory variation to its source. Mandarin Chinese lexical tones are coarticulated based on surrounding tones. We tested how L1-Mandarin and naive listeners compensate for tonal coarticulation using a paired discrimination task. L1 listeners showed greater perceptual sensitivity to tonal differences than tone-naive listeners. Yet, both L1 and tone-naive listeners showed differences in sensitivity to differently-coarticulated versions of the rising tone presented in the same tonal context. In different tonal contexts, both groups showed similar patterns of perceptual compensation for tonal coarticulation. Thus, although L1 and naive Mandarin listeners show different sensitivities to tonal variation, they display similar compensatory patterns for tonal coarticulation.

KEYWORDS

tonal coarticulation, perceptual compensation, Mandarin, lexical tones, L2 speech perception

1 Introduction

Coarticulation is a phenomenon whereby speech sounds are affected by gestural overlap with nearby speech sounds. This may occur with adjacent speech sounds, as in vowel nasalization in English syllables with a nasal coda (Zellou, 2017) - it may also occur across syllables, as in vowel-to-vowel coarticulation, where the production of a vowel may be influenced by a following vowel, even if there is an intervening consonant (Magen, 1997).

Coarticulation also occurs with lexical tones, where preceding and following tones may influence both a tone's relative pitch height and its contour. Tonal coarticulation has been examined in various languages, including Vietnamese (Brunelle, 2009), Thai (Gandour et al., 1994), and Mandarin Chinese (Shen, 1990; Xu, 1997). Tonal coarticulation in these languages shows strong assimilatory carryover effects (Gandour et al., 1994; Xu, 1997; Brunelle, 2009), while Thai and Mandarin also show some weaker dissimilatory anticipatory effects in tonal coarticulation for certain tones (Gandour et al., 1994; Xu, 1997).

Standard Mandarin Chinese is a tonal language with four tones plus one neutral tone (Cheng, 2011). Three of the four are contour tones: a rising tone (T2), a falling tone (T4), and a falling-rising tone (T3). The actual realization of these tones in connected speech is influenced by a variety of factors, including tone sandhi and tonal coarticulation. For example, T3 is usually only realized as a falling-rising tone in isolated or final contexts. T3 sandhi effects

occur where T3 becomes T2 before a second T3, and where T3 often becomes a low-falling tone in all other contexts (Chen, 2000). As for tonal coarticulation in Mandarin, tones display an assimilatory carryover effect, where the pitch height of the offset of one tone influences the pitch height of the onset of the next tone, such that a tone will have a higher onset if it is preceded by a tone with a high offset (Shen, 1990; Xu, 1997). In addition, a small dissimilatory anticipatory effect can also be seen where the maximum f0 of a tone increases when preceding a tone with low onset, and decreases when preceding a tone with high onset (Xu, 1997; Zhang and Xie, 2020). Thus, a rising tone that occurs after a low-falling tone and before a high tone will have a lower onset (assimilatory carryover effect) and a slightly lower offset/maximum f0 (dissimilatory anticipatory effect).

In general, coarticulated rising and falling tones in Mandarin share the same contour shapes as variants produced in isolation. However, in some contexts, variants may occur where the direction of the pitch contour is flattened or even slightly reversed. In a *continuous* (or "compatible") context, the offset of the preceding tone is at a similar pitch height to the onset of the next tone. For example, if a rising tone, which has a low onset and high offset, is preceded by a low-falling tone and followed by a high tone, said rising tone will be realized in its *canonical* form because the pitch onset and offset of the tonal heights in the neighboring syllables match. On the other hand, if a rising tone is preceded by a high tone and followed by a low-falling tone, this context is *discontinuous* (or "conflicting") because there is a mismatch between the pitch offset and onset of the tonal patterns at each syllable boundary and that of the target vowel. In discontinuous contexts, Mandarin contour tones are coarticulated such that the pitch contour can be realized as entirely flattened or even pushed slightly in the opposite direction. This may result in a rising tone realized with a flattened or "downward gliding" pitch, especially in fast speech (see Figure 3 in Xu, 1994).

When it comes to perception of these variants, when Xu (1994) presented these flattened tonal variants to L1 listeners in isolation, listeners performed below chance at correctly identifying the underlying tone, often identifying flattened T2 or T4 as a high tone. However, when these variants were presented in context, L1 listeners were easily able to identify the underlying tone (Xu, 1994). Thus, while tonal coarticulation leads to a great deal of variation in how, e.g., a rising tone will be realized across contexts, this context-dependent tonal variation does not reduce word intelligibility for native Mandarin listeners in connected speech.

What can account for this? On the one hand, listeners have been shown to ascribe coarticulatory effects in the signal to their sources. For instance, L1 English listeners may hear the vowels in the words "bad" and "band" as the same, even though they are phonetically different due to nasalization, because they ascribe the coarticulatory variation as stemming from the final consonant (Kawasaki, 1986). This perceived similarity of acoustically different sounds occurring in different contexts is known as *perceptual compensation for coarticulation*. Full perceptual compensation for coarticulation involves listeners neutralizing all context-related acoustic details and retaining only the invariant linguistic information (Fowler, 2006).

On the other hand, perceptual compensation for coarticulation is often only partial, with listeners perceptually attributing differences in acoustic detail to differences in context, but retaining that acoustic detail as well. For example, listeners show partial perceptual compensation for vowel nasality (Beddor and Krakow, 1999; Zellou et al., 2020). In fact, tonal coarticulatory information can indicate prosodic boundaries, and as such has been shown to be used by native listeners in speech segmentation (Lai and Kuang, 2016; Guo and Ou, 2019). Furthermore, cross-linguistic studies of segmental coarticulation effects have shown that listeners are able to use coarticulatory variation to perceive invariant linguistic information (Beddor, 2009). For example, native American English listeners look more quickly and accurately at images depicting the word "bent" when hearing coarticulatory nasalization on the vowel alone, indicating that coarticulation is not fully subtracted, but rather provides early and informative perceptual cues about upcoming segmental information (Beddor et al., 2013; Zellou, 2022). Like produced patterns of coarticulation, perceptual compensation for coarticulation is language-specific; for example, L1 Thai and L1 English speakers show different patterns of perceptual compensation for vowel nasality in American English due to differences in patterns of produced coarticulatory vowel nasality across these languages (Beddor and Krakow, 1999). And, many researchers have come to similar conclusions, that perceptual compensation for coarticulation exhibits language-specific patterns (Beddor et al., 2002; Darcy et al., 2007; Han et al., 2012).

While there is a substantial body of work on the perception of segmental coarticulation effects across languages, less is known about L1- or L2- (or, as in this study, naive-) listener perceptual compensation of tonal coarticulation. Xu (1994) found evidence for perceptual compensation for tonal coarticulation by L1 Mandarin listeners. While flattened T2 was correctly perceived as T2 when occurring in the original (discontinuous) context, flattened T2 was perceived predominantly as a high tone when presented in isolation, demonstrating the effect of context in accurate tone perception. Similarly, when flattened T2 was spliced into a continuous context (for example, T3-T2-T1), the flattened T2 was most often perceived as a falling tone. Thus, L1 Mandarin speakers use perceptual cues in the context to identify a target tone, demonstrating perceptual compensation for tonal coarticulation (Xu, 1994).

To our knowledge, no study has investigated L2 or tone-naive compensation for tonal coarticulation. In general, L2 tone perception is challenging, especially for those with no L1 tone language experience (Pelzl, 2019). However, even without L1 tone categories, learners may draw on experience from intonation and prosody in their L1. L1 English-L2 Mandarin tone acquisition has been analyzed in terms of two models of L2 phonological acquisition: the Speech Learning Model (SLM; Flege, 1995; Hao, 2014) and the Perceptual Assimilation Model for Suprasegmentals (PAM-S; So and Best, 2008, 2010). The SLM predicts acquisition of L2 sounds based on similarity to L1 sounds: sounds that are the same will be easily and quickly acquired, sounds that are different will be slow to acquire, and sounds that are similar will be initially easy but may never be fully acquired (Flege, 1995). PAM-S predicts discriminability of L2 contrasts based on assimilation of L2 sounds into L1 categories (Best, 1994). L1 English speakers do assimilate Mandarin tones into English prosodic categories (So and Best, 2008), and these assimilation patterns can predict naive listeners' ability to discriminate Mandarin tones (So and Best, 2010). However, neither the SLM nor PAM-S predicts how or whether learners will perceptually compensate for tonal coarticulation.

This gap in cross-language work on perceptual compensation of tonal coarticulation is addressed in this study, which uses a forced-choice discrimination task [as opposed to a tone identification task as in Xu (1994)] to permit comparison of perceptual compensation across groups that have differing levels of tone experience.

The present study asks the following question: How do tone-native (L1 Mandarin) and tone-naive (L1 English) listeners compensate for tonal coarticulation in Mandarin? Learning to compensate for tonal coarticulation may be challenging for tone-naive listeners, as well as for L2 learners who do not speak a tonal language. In particular, while all languages use suprasegmental features like pitch in production, there are systematic differences in how pitch is realized in tonal vs. non-tonal languages; for instance intonation pitch contours in English are more variable across and within speakers than lexical tonal contours (Michaud and Vaissière, 2015). This might mean that English listeners are less sensitive to differences in context-specific lexical tone patterns in Mandarin. Moreover, much research shows that L2 speakers of Mandarin are more accurate at tonal identification when tones appear in isolation rather than in disyllabic, trisyllabic, or sentential context (Xu, 1997; He and Wayland, 2013; Yang, 2015). Thus, it could be that part of L2 and tone-naive listeners' difficulties identifying tones in connected speech stems from the difficulty of perceptually compensating for the tonal coarticulation that occurs in multi-word contexts. Thus, the results from the current study could have applications for learning a second language that uses lexical tones.

On the other hand, tone-naive listeners have been shown to be sensitive to phonetic variation of tone contours (e.g., Hallé et al., 2004), though that study only looked at tone presented in monosyllabic contexts. Additionally, tone-naive listeners may be more sensitive than L1 listeners to small within-tone variations in contour even when the fluctuations are due to coarticulation, perhaps because L1 listeners perceive these variants as phonologically the same (Stagray and Downs, 1993). Leading models of L2 phonological acquisition [e.g., PAM-S (So and Best, 2010, 2014); and SLM (Flege, 1995)] do not account for how perceptual compensatory patterns transfer during second language learning. The current study can speak to this gap.

2 Materials and methods

2.1 Materials

The stimuli were generated using the Microsoft Azure Text-to-Speech system, in Mandarin Chinese with the Xiaoxiao (晓晓) voice at 0.9 speed. Our motivation in using a TTS voice is that our stimuli could be recreated by another lab working on similar questions. Moreover, with the recent popularity of app-based language learning systems, such as Duolingo, which utilize TTS voices to teach users a foreign language, it is important to understand how perceptual processing of tones occurs in synthesized speech. We acknowledge that use of machine-generated speech is one limitation of the current study. This point is discussed more thoroughly in the discussion.

A carrier sentence, 我告诉你_____究竟是什么 (wǒ gàosu nǐ _____ jiūjìng shì shénme; "Let me tell you what ___ is really about") was used. The target forms were different versions of the nonword *bataba* with different tones on each syllable. These segmental syllables are phonotactically allowed in both English and Mandarin but together carry no meaning in either language.

The two primary nonwords generated for the experiment were *ba*↓ *ta*↑ *ba*↑ and *ba*↑ *ta*↰ *ba*↓ (the symbol ↰ is used to represent the coarticulatorily-flattened rising tone). The first word's low-rising-high tone sequence provides a *continuous* pitch context for the middle syllable's rising tone, which thus is coarticulated to have its *canonical* rising form: *ta*↑. The second word's high-rising-low tone sequence provides a *discontinuous* context for the middle syllable's rising tone, resulting in a coarticulated *flattened* form of the rising tone: *ta*↰.

The stimuli for the perception experiment were generated by cross-splicing syllables so that a flattened rising tone on the syllable *ta*↰ was spliced into a continuous context *ba*↓_*ba*↑, where the canonical form of a rising tone would normally be expected. Similarly, a canonical rising tone on the syllable *ta*↑ was spliced into a discontinuous context *ba*↑_*ba*↓, where a flattened form of the rising tone would normally be expected.

This generated four versions of the nonword *bataba*: continuous context-canonical tone (CC form), continuous context-flattened tone (*CF* form), discontinuous context-canonical tone (DC form), and discontinuous context-flattened tone (DF form). The CC and DF forms are coarticulated *appropriately* for their context, while the *CF* and DC forms are coarticulated *inappropriately* for their context.

Two additional *bataba* wordforms with a middle falling tone were generated using the same carrier sentence and TTS engine for use in filler trials; one with high-falling-low tones (*ba*↑ *ta*↘ *ba*↓) and one with low-falling-high tones (*ba*↓ *ta*↘ *ba*↑; the symbol ↘ is used to refer to the coarticulatory-flattened falling tone). The middle syllables from these words were spliced between the same *ba*↓_*ba*↑ and *ba*↑_*ba*↓ frames as for the rising tones. This resulted in an additional two wordforms: a high-falling-low form (HFL form, *ba*↑ *ta*↘ *ba*↓) and a low-falling-high form (LFH form, *ba*↓ *ta*↘ *ba*↑).

2.2 Participants

Two participant groups - a tone-naive group (L1 English speakers) and an L1 Mandarin group - were recruited from the UC Davis psychology subjects' pool. All participants received partial course credit for their participation.

The tone-naive group consisted of 21 L1 English listeners (mean age 19.6 years, age range 18–24; 14 female, 7 male). There were 7 monolingual English speakers in this group; the other 14 reported at least one other language (Arabic, French, Hmong, Japanese, Kannada, Korean, Russian, and/or Spanish). Additional listeners who reported experience with a tone language ($n = 5$), who reported not being L1 English speakers ($n = 2$), or who did not pass the comprehension check question ($n = 1$) were excluded.

The L1 Mandarin group consisted of 23 participants (mean 19.6 age years, age range 18–25; 18 female, 5 male). All participants reported that their first language, the language they spoke at home before age 5, and their strongest language was Mandarin Chinese. The L1 Mandarin participants reported having lived in the US between 0 and 10 years with an average of 2.8 years (7 reported living in the US for less than a year; 9 reported living in the US for 1–3 years; 7 reported living in the US for 4 or more years). Additional participants who specified their first language as Chinese but did not specify Mandarin were excluded ($n = 3$), and participants who did not pass the comprehension check question ($n = 3$) were also excluded.

2.3 Procedure

The experiment was conducted online via a Qualtrics survey. Before beginning the study, participants were instructed to verify that they were in a quiet room, that all other applications and browser tabs were closed on their computer, and that their phone was on silent. Before the

experiment began, participants completed a sound calibration procedure in which they heard one sentence produced by a speaker (not used in experimental trials), presented in silence, and were asked to identify the sentence from three multiple choice options, each containing a phonologically close target word (e.g., they heard "Bill asked about the host" and were given options for sentences ending in host, toast, coast).

All participants completed a 4IAX (paired discrimination) task. On each trial, they heard two pairs of the nonce wordforms. Their task was to identify which of the two pairs contained different-sounding *middle syllables*. The trial instructions were: "Which pair of words has the more different sounding middle syllables? Pay attention only to the middle syllable."

There were four test trial types along two variables: Same and Different Context, and Same and Different Tone. Table 1 describes each trial type. For the Same-Context Same-Tone trial types, all four forms had the same coarticulatory pitch context, either continuous or discontinuous. One pair of forms was exactly the same, with the middle syllable rising tone coarticulated appropriately for the context (for example, CC form and CC form), while the other pair of forms had different middle syllables: one with a canonical rising tone and one with a flattened rising tone (for example, CC form and *CF* form). These trials thus test participants' ability to distinguish the two different coarticulated versions of the rising tone, when those tones appear in the same context.

For the Different-Context Same-Tone trial types, each pair of forms contained one form with a continuous context and one form with a discontinuous context. One pair of forms had acoustically different middle syllables, with each rising tone coarticulated appropriately for its context (CC form and DF form). The other pair of forms had acoustically identical middle syllables, one coarticulated appropriately for the context, while the second form contained a different tonal context but the exact same middle syllable, which would be inappropriate for that context (for example, the CC and DC forms). If participants consistently select the CC and DF forms as different, they are relying on the actual acoustic differences between the middle syllables to make the differentiation. On the other hand, if participants sometimes select the CC and DC forms as having the different middle syllables, despite those syllables being acoustically identical, then they are perceptually attributing acoustic differences in the middle syllables in the CC and DF forms to the difference in context – and thus displaying perceptual compensation for tonal coarticulation.

In addition, there were Different-Tone trials. The Different-Tone trial type served as a way to assess participants' ability to discriminate two entirely different lexical tones (falling and rising) instead of two differently coarticulated rising tones (canonical or flattened), as in the Same-Tone trial types. In the Same-Context Different-Tone trials, context was the same for all pairs, and one pair was exactly identical (e.g., CC and CC forms), while the other pair differed only in that one syllable was a rising tone and the other a falling tone, both coarticulated appropriately for that context (e.g., CC and LFH forms). In the Different-Context Different-Tone trials, context differed between pairs. In one pair, both forms contained the acoustically same rising tone (e.g., CC and DC forms). In the second pair, one form contained a rising tone and the other a falling tone, with both coarticulated appropriately for that context (e.g., CC and HFL forms).

Each condition was counterbalanced such that participants heard each set of word forms in all possible orders. In total, participants completed 8 total Same-Context Same-Tone trials, 16 Different-Context Same-Tone trials, 8 Same-Context Different-Tone trials, and 16 Different-Context Different-Tone trials (48 total). Trials were presented in random order.

The study was approved by the UC Davis Institutional Review Board (IRB) and subjects completed informed consent before participating.

3 Results

Responses were coded binomially for whether the listener correctly identified the pair with acoustically different tones (=1) or not (=0). These data were modeled using a mixed-effects logistic regression model, with *lme4* R package (Bates et al., 2015). The fixed effects for the model were the participants' Language Group (L1 Mandarin or L1 English) and the comparisons in each trial: Context (Same context or Different context) and Tone (Same tone or Different tone). By-trial random intercepts and by-listener random slopes for Context and Tone were also included, and sum-coding was used. The random effects structure was simplified until the model converged (Barr et al., 2013). [Retained glmer syntax: Discrimination ~ Language Group * Context * Tone + (1 | Listener) + (1 | Trial)]. The full model output is shown in Supplementary Table S1. Discrimination performance across the conditions of Language Group, Context, and Tone is shown in Figure 1.

First, the model revealed a simple effect of Context, with greater transcription accuracy for trials in the Same Context condition

TABLE 1 Examples of trials in each condition.

Trial type	Pair 1	Pair 2	Description
Same context, different tone	CC form: *ba↓ ta↗ ba↗* CC form: *ba↓ ta↗ ba↗*	CC form: *ba↓ ta↗ ba↗* LFH form: *ba↓ ta↘ ba↗*	Tests ability to discriminate between two different tones in the same context
Different context, different tone	CC form: *ba↓ ta↗ ba↗* DC form: *ba↗ ta↗ ba↓*	CC form: *ba↓ ta↗ ba↗* HFL form: *ba↗ ta↘ ba↓*	Tests ability to discriminate between two different tones when context differs within pairs
Same context, same tone	CC form: *ba↓ ta↗ ba↗* CC form: *ba↓ ta↗ ba↗*	CC form: *ba↓ ta↗ ba↗* *CF* form: *ba↓ ta↑ ba↗*	Tests ability to discriminate between two differently-coarticulated versions of one tone in the same context
Different context, same tone	CC form: *ba↓ ta↗ ba↗* DC form: *ba↗ ta↗ ba↓*	CC form: *ba↓ ta↗ ba↗* DF form: *ba↗ ta↑ ba↓*	Tests ability to discriminate between two differently-coarticulated versions of one tone in different contexts; listeners who perceptually compensate are predicted to perform around chance

The forms in the first pair have acoustically identical middle syllables.

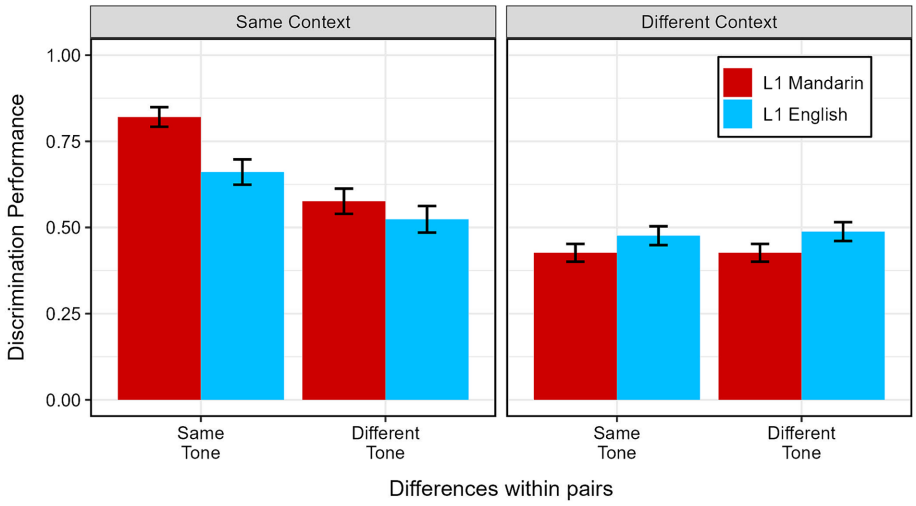

FIGURE 1
Mean discrimination performance for tones, by language group, context, and whether target tones belonged to the same tonal category. Error bars represent standard error.

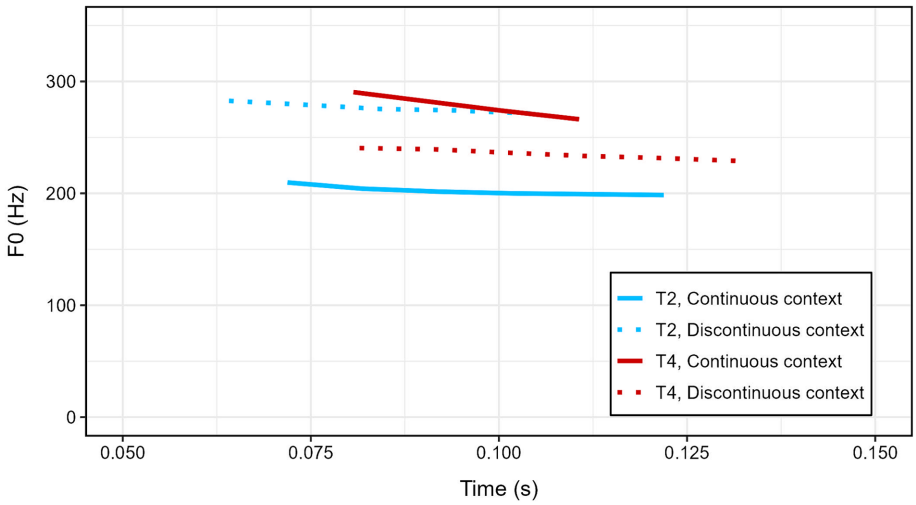

FIGURE 2
Tonal contours for the four target middle syllables of the stimuli, with rising tones in blue and falling tones in red. Solid lines denote a continuous pitch context when that tone was elicited, which predicts the "canonical" form of the rising and falling tones. Dashed lines denote a discontinuous pitch context when that syllable was elicited, which predicts a flattened or even reversed contour for rising and falling tones produced in that context.

(*Coef* = 0.42, SE = 0.05, $z = 7.84$, $p < 0.001$). There was also a simple effect for Tone, with greater transcription accuracy for trials in the Same Tone condition (*Coef* = 0.22, SE = 0.05, $z = 4.10$, $p < 0.001$). The model revealed an interaction between Language Group and Context, with greater accuracy for L1 Mandarin speakers in the Same Context condition (*Coef* = 0.19, SE = 0.05, $z = 3.86$, $p < 0.001$). There was also an interaction between Context and Tone, wherein there was greater likelihood in selecting the acoustically-different vowels in trials where the tonal context was the same within pairs, and where the middle syllables all shared the same tone (*Coef* = 0.23, SE = 0.05, $z = 4.32$, $p < 0.001$).

The findings that identification of acoustically-different syllables was higher for trials in the Same Tone condition, especially for L1 Mandarin speakers, is an unusual finding. We would expect speakers to be more likely to identify two syllables as different if they have different tones, versus two differently coarticulated versions of the same tone. This should be especially more likely for L1 Mandarin talkers. If the tone pairs in question are categorized as different tones by the L1 Mandarin listeners, it should not make a difference whether the tones appear in the same context or in a different context, as L1 Mandarin speakers must constantly categorize tones that appear in many different contexts in order to successfully perceive speech.

This unexpected finding led us to inspect the stimuli used in the experiment. Contours for the four target syllables used in the stimuli are presented in Figure 2. We found that the machine-generated tones did not have the contour patterns predicted by

Xu (1997). Instead of a rising contour, the canonical rising tone, elicited in a low-rising-high context, displays a downward-gliding contour. Meanwhile, the flattened rising tone and the canonical falling tone, which are both produced in the same high-tone-low context, have almost exactly the same tonal contour. This results in a situation where comparisons of the same rising tone have an acoustically greater difference than many of the possible comparisons between iterations of the rising tone and the falling tone. We also suspect that L1 Mandarin listeners may not have phonologically categorized the intended rising and falling tones as expected. Due to these discrepancies, we suspect that an analysis of the data based on a trial type breakdown of same versus different tones is not supported. Instead, we proceeded with an analysis based on the acoustic difference between the tones compared in each trial.

3.1 Results by acoustic distance in pitch contour

In order to explore whether the acoustic differences between different tones affected listener performance, we conducted a *post hoc* analysis of the results. As an acoustic measure of distance between tonal contours, we used the RMS (root-mean-square) of the difference between the two contours at the beginning, midpoint, and end of the contours. Table 2 shows a breakdown of the RMS for the different contrasts tested in the experiment. As seen, the means confirm what was shown in Figure 3: there are differences in how distant the tones are across conditions.

In order to test the effect of tonal distances on listener responses, the data was then modeled using a mixed-effects logistic regression model, with *lme4* R package (Bates et al., 2015). The RMS values were centered and scaled prior to model fitting. As in the previous model, the fixed effects for the model included the participants' Language Group (L1 Mandarin or L1 English) and whether context varied within pairs for each trial: Context (Same context or Different context). However, instead of Tone, we included RMS as a fixed effect. By-trial random intercepts and by-listener random slopes for Context and RMS were also included, and sum-coding was used. The random effects structure was simplified until the model converged (Barr et al., 2013). [Retained glmer syntax: Response ~ Language Group * Context * RMS + (1 + Context | Listener)]. The model output is shown in Supplementary Table S2 and mean accuracy across the conditions of Language Background, Context, and RMS is shown in Figure 3.

First, there was a significant effect for Context, where listeners are more likely to identify the acoustically-different vowel pairs in trials where pairs shared the same context (*Coef* = 0.52, SE = 0.08, $z = 6.60$, $p < 0.001$). Simple effects for Language Group (*Coef* = 0.12, SE = 0.06, $z = 1.95$,

TABLE 2 RMS of the difference in tonal contour, for each contrast tested.

Tone in Pair 1	Tone in Pair 2	RMS	Contrast appears in these conditions
Canonical rising	Flattened rising	71.6	Same Context, Same Tone Different Context, Same Tone
Canonical rising	Flattened falling	30.4	Same Context, Different Tone
Flattened rising	Canonical falling	6.5	Same Context, Different Tone
Canonical rising	Canonical falling	75.9	Different Context, Different Tone
Flattened rising	Flattened falling	41.6	Different Context, Different Tone

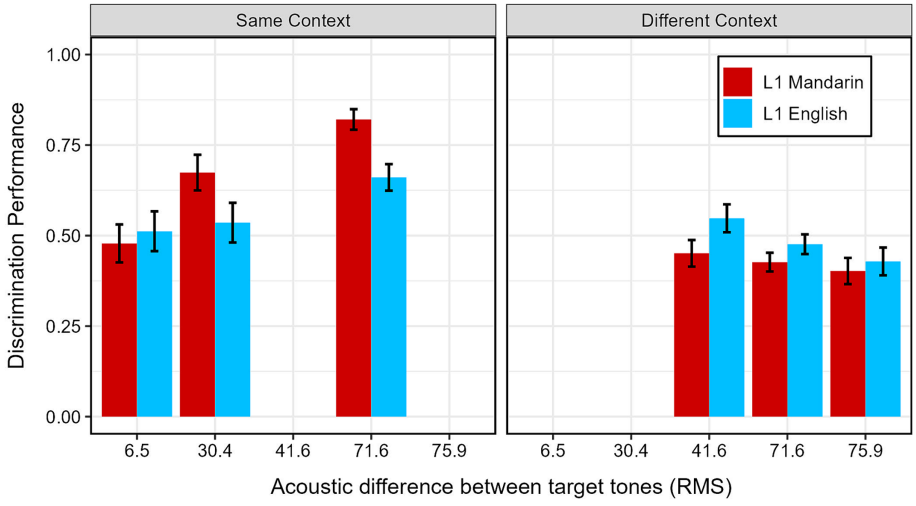

FIGURE 3
Mean discrimination performance, by language group, context, and acoustic difference between tones. Error bars represent standard deviation.

$p = 0.05$) and RMS (*Coef* = 0.10, SE = 0.05, $z = 1.86$, $p = 0.06$) were not significant.

We also see interactions involving Language Group. L1 Mandarin talkers were especially likely to discriminate acoustically-different vowels in the Same Context condition (*Coef* = 0.26, SE = 0.08, $z = 3.29$, $p < 0.01$) and were more able to benefit from the acoustic differences within pairs (*Coef* = 0.12, SE = 0.05, $z = 2.18$, $p = 0.03$).

Additionally, when context was the same, all listeners were more able to benefit from acoustic differences between pairs (*Coef* = 0.28, SE = 0.05, $z = 5.21$, $p < 0.001$).

In order to confirm effects of language background and acoustic distance in the two contexts, *post hoc* mixed-effects logistic regression models were run on subsets of the data. For the model of trials in the Same Context condition, the fixed effects included RMS (centered and scaled) and Language Group. Again, by-trial random intercepts and by-listener random slopes for Context and RMS were also included, and this random effects structure was simplified until the model converged [Barr et al., 2013; Retained glmer syntax: Discrimination ~ Language Group * RMS + (1 | Listener) + (1 | Trial)]. The full model output is shown in Supplementary Table S3.

This model confirms the effects of Language Group (*Coef* = 0.28, SE = 0.10, $z = 2.76$, $p < 0.01$) and RMS (*Coef* = 0.49, SE = 0.10, $z = 5.08$, $p < 0.001$) on the ability of participants to select the acoustically different pairs. Further, an interaction between Language Group and RMS shows that L1 Mandarin listeners are more likely than L1 English listeners to select the acoustically different syllable, when the acoustic difference between those syllables is greater (*Coef* = 0.20, SE = 0.08, $z = 2.41$, $p = 0.02$).

For the *post hoc* model of trials in the Different Context condition, the same fixed effects and random effects structure was used, and the random effects structure was simplified until the model converged [Barr et al., 2013; Retained glmer syntax: Discrimination ~ Language Group * RMS + (1 | Listener)]. The full model output is shown in Supplementary Table S4.

This model shows that when context differs within pairs, there is an effect of RMS on listener identification of the acoustically different syllables, where listeners are actually less likely to choose the acoustically different syllables as the different pair, when the acoustically different syllables are more different (*Coef* = −0.11, SE = 0.05, $z = -2.10$, $p = 0.04$), the opposite directionality as in the Same Context condition. There was no effect of Language Group for trials in the Different Context condition (*Coef* = −0.12, SE = 0.08, $z = -1.49$, $p = 0.14$).

4 Discussion

The goal of the current study was to investigate L1 Mandarin and tone-naive (L1 English) listeners' perceptual compensation for tonal coarticulation, as well as tone-naive listeners' perceptual sensitivity to acoustic differences in tones. Overall, tone-naive listeners were able to perceive the differences between syllables with acoustically different pitch levels, although to a lesser extent than that displayed by L1 Mandarin listeners. Yet, both groups showed similar patterns of perceptual compensation on the Different-Context trials, attributing differences in the acoustics to differences in the tonal contexts. In the following paragraphs, we break down each of the key findings from the current study and discuss their relevance for understanding cross-language speech perception, as well as implications for learning of a tonal language.

Our first key finding is that, in the same-context conditions, there were differences in tone perception across L1 and naive listeners. L1 listeners showed greater performance in identifying different lexical tones than naive listeners when target tones occurred in the same tonal context. This is consistent with the fact that native language tonal experience provides an advantage in tone discrimination (Hao, 2012; Yang, 2015; Lee and Lee, 2022). However, for both listener groups, identification of different tones in the same context was modulated by degree of acoustic difference – i.e., tone-naive listeners were still sensitive to acoustic tonal differences when context did not differ within pairs. This suggests that greater acoustic differences across tones can be leveraged by tone-naive listeners as well (as long as target words occur in the same phonological context).

Secondly, in contrast to the same-context conditions, when tonal contexts differed within pairs, overall performance at identifying the acoustic differences than in the same-context conditions was similar for L1 and naive listeners. The lower performance for different-context pairs is expected because compensation for tonal coarticulation should make veridical perception of tonal differences harder. In other words, listeners are compensating for tonal coarticulation, making tonal perception difficult. What is surprising is that this does not vary with language background. L1 and naive listeners appear to be compensating similarly. This is in contrast with much prior work using identical methods to ours that do find differences across language background for compensation for segmental coarticulation effects (Beddor and Krakow, 1999; Beddor et al., 2002). Why is it the case that language experience does not lead to differences in compensation for tonal coarticulation, but other studies have found language-specific differences in compensation for segmental coarticulation? One possibility is that even though lexical tone-naive (here, L1 English) listeners do not have phonological experience with tone, they do have lots of experience with prosody and tonal contours indicating pragmatic articulations in their native language (e.g., marking questions vs.statements, contrastive focus marking, etc.). Perhaps the pragmatic uses of pitch leads listeners to compensate for lexical tone in the same way as phonological-tone listeners. Another possibility is that the variation in the machine-generated stimuli in this study does not match with actual coarticulatory variation in Mandarin, such that L1 Mandarin listeners do not experience a benefit of experience.

Finally, we observe that gradient acoustic variations across our items led to differences in tone discrimination. In particular, in same-context conditions, listeners showed greater perceptual sensitivity when acoustic difference between tones was greater, and L1 Mandarin listeners showed even greater perceptual sensitivity than tone-naive listeners. This is another case where we observed that native language experience matters in tone perception: greater perceptual sensitivity of larger acoustic differences in tones. On the other hand, we do not find an effect of language experience for acoustic differences in different contexts. Again, this supports our interpretation that phonological experience with tones provides a benefit in tonal discrimination when the surrounding phonological context is the same, but not when the context contains varying tonal patterns (where compensation makes tone perception difficult for all listeners). In other words, just like the perception of pitch in isolation, tone perception in the same contexts is easier for L1 listeners because they have experience attending to pitch for lexical identification; yet, in the different context, compensation is triggered. This could indicate that compensation for pitch is not dependent on phonological experience with tone.

The finding that compensation for coarticulation in a second language works differently for segmental coarticulatory patterns (e.g., Beddor and Krakow, 1999) and tonal coarticulatory patterns (current study) could also be used to incorporate the role of coarticulatory perception in L2 phonological acquisition frameworks [e.g., PAM-S (So and Best, 2010, 2014); and SLM (Flege, 1995)]. For example, in the PAM-S framework, discrimination of L2 tones is predicted by assimilation of those sounds to particular L1 categories. However, the present study demonstrates that the surrounding context also influences L2 listeners' ability to discriminate L2 tones. Incorporating the effect of context in perception in these models is a ripe avenue for future work.

In general, our findings are consistent with prior work reporting the facilitative effects of tonal experience and the impact of acoustic similarity on the tone perception (Zhu et al., 2021, 2023). Above and beyond the implications for the processing of lexical tones for listeners of different language backgrounds, the findings from this study can also be informative to understanding the acquisition of lexical tones by adult language learners. In particular, since we found the largest difference across listener groups for tone perception in same tonal contexts, it could be most helpful in pedagogical or educational contexts to target those types of phrases for learners. As mentioned in the Introduction, examining how learners learn to perceive tones in multi-phrase utterances is relatively understudied (Chang and Bowles, 2015; Hao, 2018). The findings in the current study indicate that this can be a ripe direction for future work to understand the particular benefits or challenges that different phonological contexts can provide to a tone-naive learner of Mandarin. Learning to perceive tone in isolation is one skill, but sensitivity to different tones in multi-word utterances is another important part of learning a tonal language like Mandarin. More work exploring perception, processing, and acquisition of tones in context is critical to understanding how adults learn tone languages.

It is worth noting that there are many individual-differences factors that might influence a listener's ability to perceive pitch differences across contexts. For instance, musical experience has long been observed to affect pitch and speech perception (e.g., Wayland et al., 2010; Cohn et al., 2023). In the current study, we did not collect information on listeners' musical training experience. Exploring factors such as this in future work can further illuminate the factors that might make learners better at perceiving tonal languages across different phonological contexts. We also note that many of the tone-naive listeners in the present study were bilingual. The role that different language experiences have on the perception of L2 tonal patterns is a direction for future work. The current study was also conducted online, so different listeners could have had different listening equipment and situations. On the other hand, perception studies comparing online and in-person modalities tend to find the same results online as in person (Woods et al., 2015). Additionally, the sound calibration procedure and participants' verifications that they were in a quiet listening environment should serve to minimize the impact of the variability in listening environments on the study.

One major limitation of this study was the use of machine-generated stimuli. We did not formally test the naturalness or intelligibility of these items used in this study. One direction for future work is to examine the effect of naturalness of TTS Mandarin on perceptual compensation and identification across listener groups. Future studies can also use human voices, or carefully manipulate computer voices, in order to investigate perceptual compensation of tones and unravel the contributions of phonetic and phonological speech perception to tone discrimination. Other future directions include examining tones other than Mandarin rising and falling tone, or tones in languages other than Mandarin. Finally, an investigation into L2 Mandarin learners' perceptual sensitivity to coarticulatory variation in same and different contexts is a step to furthering our understanding of L2 tonal perception and the time course of the acquisition of a novel phonological dimension.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://doi.org/10.17605/OSF.IO/2WPZS.

Ethics statement

The studies involving humans were approved by UC Davis Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JV: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. GZ: Conceptualization, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2024.1392022/full#supplementary-material

References

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Machler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beddor, P. S. (2009). A Coarticulatory path to sound change. *Language* 85, 785–821. doi: 10.1353/lan.0.0165

Beddor, P. S., and Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: evidence for partial compensation. *J. Acoust. Soc. Am.* 106, 2868–2887. doi: 10.1121/1.428111

Beddor, P. S., Harnsberger, J. D., and Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *J. Phon.* 30, 591–627. doi: 10.1006/jpho.2002.0177

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., and Brasher, A. (2013). The time course of perception of coarticulation. *J. Acoust. Soc. Am.* 133, 2350–2366. doi: 10.1121/1.4794366

Best, C. T. (1994). "The emergence of native-language phonological influences in infants: A perceptual assimilation model" in *The Development of speech perception: the transition from speech sounds to spoken words* (Cambridge, MA: MIT Press), 167–224.

Brunelle, M. (2009). Northern and southern Vietnamese tone coarticulation: a comparative case study. *J. Southeast Asian Linguist.* 1, 49–62.

Chang, C. B., and Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *J. Acoust. Soc. Am.* 138, 3703–3716. doi: 10.1121/1.4937612

Chen, M. Y. (2000). Tone sandhi: Patterns across Chinese dialects. Cambridge, United Kingdom: Cambridge University Press.

Cheng, C. C. (2011). A synchronic phonology of Mandarin Chinese. The Hague, Netherlands: De Gruyter Mouton.

Cohn, M., Barreda, S., and Zellou, G. (2023). Differences in a musician's advantage for speech-in-speech perception based on age and task. *J. Speech Lang. Hear. Res.* 66, 545–564. doi: 10.1044/2022_JSLHR-22-00259

Darcy, I., Peperkamp, S., and Dupoux, E. (2007). Bilinguals play by the rules: perceptual compensation for assimilation in late L2-learners. *Lab. Phonol.* 9, 411–442.

Flege, J. E. (1995). "Second language speech learning: theory, findings, and problems" in Speech perception and linguistic experience: Issues in cross-language research (Timonium, MD: York Press).

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Percept. Psychophys.* 68, 161–177. doi: 10.3758/BF03193666

Gandour, J., Potisuk, S., and Dechongkit, S. (1994). Tonal coarticulation in Thai. *J. Phon.* 22, 477–492. doi: 10.1016/S0095-4470(19)30296-7

Guo, Z., and Ou, S. (2019). "The use of tonal coarticulation in speech segmentation by listeners of Mandarin." in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Available at: https://assta.org/proceedings/ICPhS2019Microsite/pdf/full-paper_138.pdf. (Accessed November 7, 2024).

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

Han, J. I., Choi, T. H., and Choi, Y. M. (2012). Language specificity in perceptual compensation for native and non-native assimilation. 언어 37, 445–480.

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

Hao, Y.-C. (2014). "The application of the speech learning model to the L2 Acquisition of Mandarin Tones." in *Proceedings of the 4th International Symposium on Tonal Aspects of Languages (TAL 2014), 67–70*. Nijmegen.

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

He, Y., and Wayland, R. (2013). Identification of Mandarin coarticulated tones by inexperienced and experienced English learners of Mandarin. *Chin. Sec. Lang. Res.* 2, 1–21. doi: 10.1515/caslar-2013-0020

Kawasaki, H. (1986). Phonetic explanation for phonological universals: the case of distinctive vowel nasalization. Experimental phonology, 81–103.

Lai, W., and Kuang, J. (2016). Prosodic grouping in Chinese trisyllabic structures by multiple cues–tone coarticulation, tone sandhi and consonant lenition. *Proc. Tonal Aspects Lang.* 2016, 157–161. doi: 10.21437/TAL.2016-34

Lee, K., and Lee, O. J. (2022). Native and non-native perception of Mandarin level tones. *Linguistic Res* 39, 567–601. doi: 10.17250/khisli.39.3.202212.007

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *J. Phon.* 25, 187–205. doi: 10.1006/jpho.1996.0041

Michaud, A., and Vaissière, J. (2015). Tone and intonation: introductory notes and practical recommendations. *KALIPHO-Kieler Arbeiten zur Linguistik und Phonetik* 3, 43–80.

Pelzl, E. (2019). What makes second language perception of Mandarin tones hard? A non-technical review of evidence from psycholinguistic research. *CSL*. 54, 51–78. doi: 10.1075/csl.18009.pel

Shen, X. S. (1990). Tonal coarticulation in Mandarin. *J. Phon.* 18, 281–295. doi: 10.1016/S0095-4470(19)30394-8

So, C. K., and Best, C. T. (2008). "Do English speakers assimilate Mandarin tones to English prosodic categories?" in *Interspeech 2008: Proceedings of the 9th Annual Conference of the International Speech Communication Association, (Brisbane, Qld.)*. p. 1120.

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

So, C. K., and Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Stud. Second. Lang. Acquis.* 36, 195–221. doi: 10.1017/S0272263114000047

Stagray, J. R., and Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language / 使用声调语言和非声调语言为母语的人对声音频率的分辨能力. *J. Chin. Linguist.* 21, 143–163.

Wayland, R., Herrera, E., and Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *J. Phon.* 38, 654–662. doi: 10.1016/j.wocn.2010.10.001

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., and Spence, C. (2015). Conducting perception research over the internet: a tutorial review. *PeerJ* 3:e1058. doi: 10.7717/peerj.1058

Xu, Y. (1994). Production and perception of Coarticulated tones. *J. Acoust. Soc. Am.* 95, 2240–2253. doi: 10.1121/1.408684

Xu, Y. (1997). Contextual tonal variations in Mandarin. *J. Phon.* 25, 61–83. doi: 10.1006/jpho.1996.0034

Yang, B. (2015). Perception and production of Mandarin tones by native speakers and L2 learners. Berlin, Heidelberg: Springer Berlin Heidelberg.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *J. Phon.* 61, 13–29. doi: 10.1016/j.wocn.2016.12.002

Zellou, G. (2022). Coarticulation in phonology. Cambridge: Cambridge University Press.

Zellou, G., Barreda, S., and Ferenc Segedin, B. (2020). Partial perceptual compensation for nasal coarticulation is robust to fundamental frequency variation. *J. Acoust. Soc. Am.* 147:EL271–EL276. doi: 10.1121/10.0000951

Zhang, H., and Xie, Y. (2020). Coarticulation effects of contour tones in second language Chinese. *Chin. Sec. Lang. Res.* 9, 1–30. doi: 10.1515/caslar-2020-0001

Zhu, M., Chen, X., and Yang, Y. (2021). The effects of native prosodic system and segmental context on Cantonese tone perception by Mandarin and Japanese listeners. *J. Acoust. Soc. Am.* 149, 4214–4227. doi: 10.1121/10.0005274

Zhu, M., Chen, F., Chen, X., and Yang, Y. (2023). The more the better? Effects of L1 tonal density and typology on the perception of non-native tones. *PLoS One* 18:e0291828. doi: 10.1371/journal.pone.0291828