

Native language phonological and phonetic influences on second language perception of Vietnamese tones

Second Language Research

1–24

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/02676583251336578

journals.sagepub.com/home/slr



Juqiang Chen 

School of Foreign Languages, Shanghai Jiao Tong University, China

Ailing Qin

School of Foreign Languages and Literatures, Guangxi University, China

Hui Chang

School of Foreign Languages, Shanghai Jiao Tong University, China

Abstract

This study examines first language (L1) phonological and phonetic influences on the perception of Vietnamese tones by Mandarin speakers learning Vietnamese as a second language (L2) in light of the Perceptual Assimilation Model. Participants were divided into two groups according to their length of classroom instruction, i.e. beginning (two months) and experienced (two years) learners. They first discriminated 15 Vietnamese tone contrasts and then mapped six Vietnamese tones onto their L1 Mandarin tone categories. The manner in which a Vietnamese tone contrast was mapped onto L1 categories (i.e. Categorized or Uncategorized) and the perceived overlap of L1 response categories of the contrast generally predicted its discrimination accuracy. Beginning and experienced learners exhibited differences in their perceptual assimilation of two Vietnamese tones but there was no significant difference in their accuracy of discriminating tone contrasts. The present study generally supports the Perceptual Assimilation Model and implications for L2 tone perception are discussed.

Keywords

lexical tones, Mandarin, second language speech learning, tone perception, Vietnamese

Corresponding author:

Ailing Qin, School of Foreign Languages and Literatures, Guangxi University, Xixiangtang District, Nanning, Guangxi 530004, P.R.China.

Email: qinailing@gxu.edu.cn

I Introduction

Native or first language (L1) experience affects non-native perception of consonants (Best et al., 2001), vowels (Tyler et al., 2014), and lexical tones (Burnham et al., 2015; Chen et al., 2020, 2023), rendering some non-native contrasts more difficult to perceive than others. In addition, increased experience with a second language (L2), i.e. the amount of learning time and exposure, improves perception of certain L2 phonetic categories/contrasts but not all of them (Hao, 2017). In the present study, we examined how L1 phonological and phonetic influences as revealed in cross-language assimilation affected discrimination of Vietnamese tones by two groups of Mandarin-native learners with different lengths of Vietnamese L2 classroom learning experience (2 months versus 2 years).

I Effects of L1 phonological and phonetic factors on non-native perception

The L1 effects on non-native speech perception are often examined through perceptual assimilation, in which listeners perceive non-native speech with reference to the phonetic categories of their native language. The Perceptual Assimilation Model (PAM: Best, 1995; PAM-L2: Best and Tyler, 2007), built on the meta-theoretical basis of ecological direct-realism and the principles of Articulatory Phonology (Browman and Goldstein, 1989), elaborates on how native phonological and phonetic (gestural and/or acoustic) properties affect perceptual assimilation of non-native phones into native categories, and how different assimilation patterns account for variations in discrimination of non-native consonant, vowel and tone contrasts (Best et al., 2001; Chen et al., 2023; Faris et al., 2018). PAM has also been extended to account for perception of tones by non-tonal native listeners, such as English and French (PAM-S: So and Best, 2011, 2014). Both PAM and PAM-S are theoretically coherent. Nevertheless, we maintain that perception of non-native tones by *tonal* native listeners should resemble that of non-native vowels and consonants because lexical tones are phonologically contrastive in their L1 system. Thus, in the present study, the original PAM was selected as the main theoretical framework and L2 intonation learning models, such as the L2 Intonation Learning theory (LILt: Mennen, 2015), were not considered. In addition, theoretical models, such as the Speech Learning Model (SLM: Flege, 1995; SLM-r: Flege and Bohn, 2021) and the Native Language Magnet model (NLM: Kuhl and Iverson, 1995), focus more on individual phonetic categories rather than on phonological contrasts. While the Second Language Linguistic Perception model (L2LP: Escudero, 2005) considers both individual categories and contrasts, it mainly concentrates on perception of vowels with predictions based on acoustic similarities. Wherever relevant, SLM and L2LP will be integrated into the discussion alongside PAM to enhance its theoretical depth.

According to PAM (Best, 1995), a non-native phone can be (1) Categorized, i.e. perceived as a good or poor exemplar of a native phoneme, or (2) Uncategorized, i.e. not perceived as any single native phoneme but still as a set of native phonemes within the native phonological space, or (3) Non-Assimilated, i.e. perceived as a non-speech sound.

In addition, when two non-native phones are Categorized¹ into contrasting native categories, i.e. Two-Category (TC) assimilations, they are predicted to be better discriminated than when two non-native phones are Categorized into a single native category with differing degrees of category goodness, i.e. Category–Goodness (CG) assimilations. When two non-native phones are Categorized as equally good or bad exemplars of a single native category, i.e. Single-Category assimilations (SC), the contrast is the most difficult to discriminate.

When a non-native contrast involves a phone within a native category and another outside the native category, i.e. an Uncategorized–Categorized (UC) contrast, it should be discriminated nearly as good as a Two-Category contrast because it is facilitated by the distinction across native phonological boundaries. When both non-native phones of the contrast fall outside the native category, i.e. an Uncategorized–Uncategorized (UU) assimilation, their discrimination ranges from being poor to good depending on their proximity to each other and to native categories (Best, 1995). These assimilation types represent different amount of L1 *phonological* constraints.

In addition, recent studies have indicated that discrimination accuracy is modulated by the degree of perceived *phonological* overlap between the two non-native phones (Antoniou et al., 2013; Best et al., 2019; Chen et al., 2023; Faris et al., 2018; So and Best, 2014). Within each contrast type above (e.g. Two-Category, Uncategorized–Categorized, Uncategorized–Uncategorized), when the contrasting non-native phones are each identified with completely different sets of native categories, i.e. *non-overlap* contrasts, they should be discriminated better than when there are one or more shared above-chance native categories, i.e. *partial-overlap* contrasts. When all the above-chance categories are the same for both non-native contrasts, i.e. *complete-overlap* contrasts, they should be the most difficult to perceive.

In addition to these phonological constraints, listeners may retain some residual *phonetic* sensitivity to differences of the non-native phones from the L1 categories. More phonetic sensitivity will facilitate better discrimination of non-native contrasts when their phonological assimilation and overlap types are the same. Thus, phonetic sensitivity measures, such as *overlap scores* (as in Flege and MacKay, 2004; Levy, 2009, i.e. the summed percentages of all overlapping native category choices, including below-chance choices) or the *fit-index* (Wu et al., 2014, by multiplying percent choice by goodness ratings) can be used to fine-tune predictions about discrimination accuracies of non-native contrasts. For example, phonetic sensitivity as indicated by overlap scores successfully predicted Thai-naïve Vietnamese-native listeners' more accurate discrimination of the mid-level and rising than the falling and low-level Thai tone contrast, even though both contrasts were assimilated as Two-Category with no phonological overlap (Chen et al., 2023).

It should be noted that phonological and phonetic factors could be defined and operationalized differently in different paradigms. For example, Yu et al. (2019) used the categorical perception paradigm in which listeners were required to perceptually map a synthetic tonal continuum into distinct tonal categories. In their study, phonological and phonetic information in lexical tones was dissociated by controlling (within-category and across-category) acoustic intervals in the stimuli. Although acoustic differences within and across native categories contribute to phonological and phonetic factors in L2

perception, ‘cross-language dissimilarity must be assessed perceptually rather than acoustically because acoustic measures sometimes diverge from what listeners perceive’ (Flege and Bohn, 2021).

Given the importance of perceptual (dis)similarity and the convention of PAM literature, our research defines *phonological* factors as indicated by perceptual assimilation and overlap types and *phonetic* factors as indicated by sensitivity scores derived from percent choice and/or goodness ratings.

Most L2 tone perception studies have tested L2 tone learners with a non-tonal L1, such as English, French or Dutch (Braun and Johnson, 2011; Braun et al., 2014; So and Best, 2011, 2014) and have attempted to attribute differences/difficulties in tone-specific perception/processing to L1 prosodic categories. However, these studies cannot fully address the issue of L1 phonological and phonetic influences on L2 tone perception because these listeners do not have phonological categories for lexical tones in their L1 and are likely to differ qualitatively from L2 learners with tonal L1s (Best, 2019; Laméris and Post, 2023; Yu et al., 2019). On the other hand, research that tested native tone language listeners with no L2 experience (e.g. Burnham et al., 2015; Chen et al., 2020, 2023; Reid et al., 2015) cannot reveal the interplay between L1 effects and L2 experience. Thus, it remains unresolved how L1 experience and varying L2 experience affect perception of L2 tones, particularly by listeners with tonal L1s.

2 Effects of L2 learning experience on L2 speech perception

The original PAM focuses on naive listeners’ performance when they perceive an unfamiliar language for the first time. As their L2 learning experience accumulates, they gradually develop a common L1–L2 system as depicted by PAM-L2 (Best and Tyler, 2007). This perceptual development can be affected by L1–L2 (dis)similarities as well as other factors, such as L2 vocabulary size (Bundgaard-Nielsen et al., 2011a, 2011b, 2012) and the frequency of the L2 category or L2 syllable in vocabulary.

Specifically, according to PAM-L2, for Two-Category and Uncategorized–Categorized assimilation contrasts, no further perceptual learning will occur for the Categorized L2 phones. Discrimination of both types of contrasts should be relatively accurate. For the Uncategorized–Uncategorized L2 contrast, if they are assimilated to different sets of L1 phones, then two new L2 phonological categories will be formed. However, if they are perceived as similar to the same set of L1 categories, a single new phonological category that encompasses the two L2 phones will emerge.

For the Category–Goodness contrast, a new category will be formed for the L2 phone that is perceived phonetically deviant from the native category whereas no further perceptual attunement will occur for the L2 phone which is perceived as a good exemplar of the native category. The most difficult contrast to perceive and to learn is the Single-Category contrast, both phones of which are perceived as phonologically and phonetically equivalent to a single native category. Whether learners can improve their distinction of this type of contrast depends on how important such contrast is at the phonological and/or lexical (i.e. the number of high frequency words that contain this contrast) levels.

Previous L2 lexical tone learning studies have examined tone learning difficulties at both sublexical and lexical levels. L2 learners have to perceive and distinguish phonetic

tone patterns and encode them as abstract phonological tone categories and then use these categories in lexical learning and processing. The present study focuses on perception of tones at the sublexical level. While there may be a continuity between sublexical and lexical processing (Braun et al., 2014; Chan and Leung, 2020; Laméris and Post, 2023; Pelzl et al., 2022; Qin et al., 2019; Wong and Perrachione, 2007), it is important to note that findings at the sublexical level cannot be directly generalized to lexical levels.

Research that investigates L2 tone perception at sublexical level often compared non-tonal L1 learners of different proficiency levels (Hao, 2017; Tsukada and Han, 2019; Tsukada et al., 2016) or compared learners with tonal and non-tonal L1s (Hao, 2012). The former group of studies find that experienced or advanced learners are better at perceiving some L2 tones and are more native-like than those with less L2 experience or lower proficiency. For example, Hao (2017) examined discrimination of Mandarin vowels and tones by L1 English learners of L2 Mandarin at different learning stages. Both beginning and experienced learners perceived /li-ly/ and Tone 1 (level tone) – Tone 4 (falling tone) contrasts more accurately than L1 English Mandarin-naïve listeners. In addition, naïve listeners were more accurate in discriminating vowel contrasts than tone contrasts which are absent in their native phonological system, whereas Mandarin experience increased learners' sensitivity to tonal distinctions, resulting in their equally good discrimination of both vowel and tone contrasts. However, there were no differences between beginning and experienced learners in discriminating both Mandarin vowel contrasts, i.e. /li-ly/, /lu-ly/ and tone contrasts, i.e. Tone 1 – Tone 4 and Tone 2 (rising tone) – Tone 3 (rising-falling tone) contrasts. The author attributed this to the fact that the beginning learner group had had more than one year experience with Mandarin and thus had already developed the ability to distinguish basic tone contrasts.

On the other hand, those studies that compared L1 tonal (e.g. Cantonese) and non-tonal (e.g. English) learners have found that perception and learning difficulties can be caused by either absence of a functional tone tier in the native phonological system (as in the case of non-tonal L1 learners) or presence of specific native tone categories that interfere with L2 tone development (as in the case of tonal L1 learners). For example, Hao (2012) tested identification of Mandarin tones by native Cantonese and English learners. Both groups had problems in distinguishing Mandarin Tone 2 and Tone 3. Additionally, the Cantonese group had difficulties in distinguishing Mandarin Tone 1 and Tone 4, which were assimilated into overlapping Cantonese tone categories, indicating native language influences on L2 tone perception. But the study did not compare Cantonese speakers with differing Mandarin learning experience and thus it remains unknown how native influences are modulated by learning experience.

To examine how L2 learning experience (i.e. time and exposure) modulates native phonological and phonetic factors in lexical tone perception, it is desirable to test native tone language learners of a new tone language via both an assimilation task and discrimination task and compare performance of learners with differing learning experience within a single study. However, most studies so far tested L2 tone discrimination and/or identification (Chang et al., 2017) but not assimilation, except for two (Hao, 2012; Wu et al., 2014). It was found that naïve listeners assimilated L2 tones to L1 tones at a phonetic level based on acoustic similarities, e.g. F0 height and contour, whereas the experienced listeners were influenced by both phonological and phonetic factors (Wu et al.,

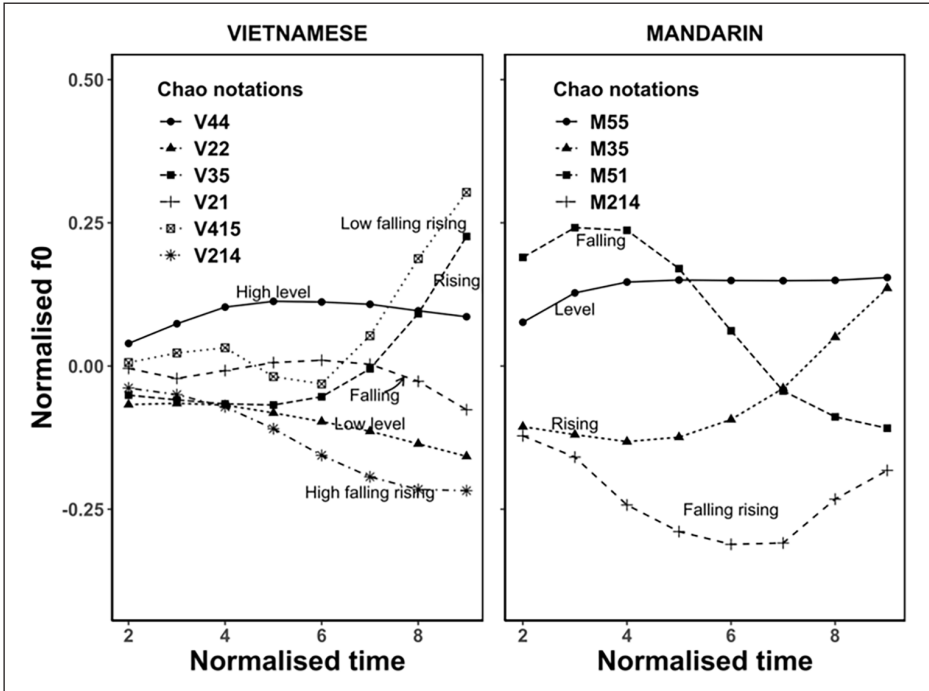


Figure 1. Time- and Lobanov-normalized (Lobanov, 1971) f_0 contours of Vietnamese and Mandarin.

Notes. For details of data used in plotting the contours, see Chen et al., 2020, 2023. The legends in each panel show the Chao notations for the tones of each language.

2014). However, this study did not include a test of discrimination. The present study addressed this issue by testing perceptual assimilation and discrimination of Vietnamese tones by L1 Mandarin learners with differing Vietnamese learning experience.

3 Lexical tones in Vietnamese and Mandarin

Following previous studies (Chen et al., 2020, 2023), a parsimonious phonological feature system was employed which includes tone contour and height features. We specify contour features as level (flat) and dynamic (i.e. rising, falling, falling–rising and rising–falling). The height specifications are high, mid, low. To minimize the dependence on specific, detailed phonetic realizations, height is specified only when it is phonologically contrastive for tones of the same contour, e.g. the height contrast for falling–rising contours in Vietnamese.

On the other hand, existing Chao transcriptions/numbers (Chao, 1930) were used to provide a phonetic approximation of the tones (Yip, 2002). Chao transcriptions are based on perceived pitch, in which 5 represents the highest and 1 the lowest pitch in speakers' vocal range. Chao transcriptions are somewhat more abstract than actual acoustic f_0 patterns in Figure 1, but not as abstract as phonological features, i.e. perceived abstract pitch

heights and *contours*. Thus, the Chao numbers sometimes differ from both the f0 contours and the phonological features of tones.

Northern Vietnamese (V), i.e. the standard dialect, has six tones in its phonological system (Brunelle, 2009): two phonologically level tones, high-level V44 (or *ngang* in Vietnamese), and low-level V22 (*huyền*); and four phonologically contour tones: rising V35 (*sắc*), falling V21 (*nặng*) and two falling–rising tones, high falling–rising² V214 (*hỏi*), and low falling–rising V415 (*ngã*) (Nhàn, 1984; Yip, 2002). Mandarin (M) has four tones (Chao, 1968), of which one is phonologically level and three are phonologically dynamic in their citation forms: a high-level tone M55 (called Tone 1 in Mandarin), a rising tone M35 (Tone 2), a falling–rising tone M214 (Tone 3), and a falling tone M51 (Tone 4) (see Figure 1).

In the present study, we report two experiments that examined phonological and phonetic factors in perception of Vietnamese tones by two groups of Mandarin-native learners with different lengths of Vietnamese L2 learning (2 months versus 2 years). The first experiment used a perceptual assimilation task to estimate L1 phonological and phonetic influences and to provide predictions about discrimination performance to be tested in experiment 2. In experiment 2, a discrimination task was used to examine if/how native language influences as indicated in perceptual assimilation patterns affected L2 discrimination.

II Experiment I

In Experiment 1, we examined how Mandarin-native learners with differing Vietnamese learning experiences assimilated Vietnamese lexical tones into their native tone system. To avoid the assimilation-task-induced L1 influence on the discrimination task, the discrimination task was run first. But for ease of reading, we report the assimilation experiment first because it was used to determine the assimilation types for the 15 Vietnamese tone contrasts and to form PAM-based predictions about discrimination.

I Method

a Participants. Two groups of Mandarin-L1 Vietnamese-L2 learners (beginning learners: $n=24$, $M_{age}=18.5$ years, $SD_{age}=0.66$, 20 females and 4 males; experienced learners: $n=25$, $M_{age}=20.4$ years, $SD_{age}=0.96$ years, 19 females and 6 males) participated in both experiments. They differed in their Vietnamese classroom learning time and exposure, which was defined as L2 learning experience in the present study. The beginning learner group consisted of first-year college students who majored in Vietnamese and learned Vietnamese for just two months in a classroom setting with no exposure to Vietnamese before college. The experienced learner group consisted of third-year college students who also majored in Vietnamese. Participants also self-rated their Vietnamese proficiency in terms of reading, speaking, listening, and writing, using a scale ranging from 1 to 100, and estimated their Vietnamese vocabulary size. Using *t*-tests, it was revealed that the beginning group reported significantly lower proficiency levels for all four language skills and a smaller vocabulary size compared to the experienced group (see Table 1).

Table 1. Self-rated Vietnamese proficiency and vocabulary size.

Proficiency	Beginner group	Experienced group	<i>T</i>	<i>p</i>
Listening	34.0	51.5	166.0	0.007
Reading	31.1	62.1	91.5	< 0.001
Speaking	34.1	52.4	167.0	0.007
Writing	25.1	57.7	93.5	< 0.001
Vocabulary	500	1400	48.0	< 0.001

The participants were born and raised in various regions in China (i.e. Guangxi, Hunan, Yunnan, Zhejiang, Hubei, Guizhou)³ but were all educated in Mandarin from early childhood through high school, and they used Mandarin on a daily basis. None had more than two years of formal musical training. All reported normal hearing. All participants gave informed consent prior to testing. The experiments were approved by the Human Research Ethics Committee of School of Foreign Languages, Shanghai Jiao Tong University (approval 2111S1102).

b Stimulus materials. Two consonant–vowel syllables (/ma/ and /mi/) recorded in a previous study (Chen et al., 2020) were used here with permission. The two syllables were selected because they form meaningful free or bound morphemes in Vietnamese and Mandarin. Two female native Northern Vietnamese informants who were born and raised in Hanoi ($M_{age}=21.5$ years) each produced 96 tokens in total (2 syllables $\times$ 6 tones $\times$ 8 repetitions) in isolation. Intensity of all auditory stimuli was normalized at 70 dB via *Praat* (Boersma, 2001).

A perceptual evaluation experiment was run to select the tokens with the best quality. Two other female native Northern Vietnamese speakers ($M_{age}=20.5$ years), who were born and raised in Hanoi and Hai Duong respectively, were presented with these tokens and were asked to identify the tones and rate the quality of pronunciation (1=poor, 7=excellent). 48 tokens, i.e. 6 tones $\times$ 2 informants $\times$ 2 vowels (/a/ and /i/) $\times$ 2 tokens, that were identified correctly by both native judges and with the highest averaged goodness ratings were selected as the stimuli.

c Procedure. Due to Covid-19 social distancing policy at the time of the experiments, participants were tested online via *Psychopy* (version 2021.1.3, Peirce et al., 2019) and were asked to do the test in a quiet environment with headphones. Before the test session, participants completed 12 practice trials to familiarize themselves with the task. The practice trials were produced by a third informant which were not used in the test session. No feedback was given in either the practice or test session.

On each trial, a stimulus token of Vietnamese tone was presented, and listeners first made a forced-choice assimilation judgment to their native Mandarin tones in four Pinyin options via mouse clicks as quickly as possible. Immediately after their assimilation response, they rated category goodness of the token as compared to their chosen native category on a 7-point scale via mouse clicks: 1=poor, 7=perfect. Each participant categorized and rated 144 trials in total, i.e. 2 talkers $\times$ 2 syllables (/ma/ and /mi/) $\times$ 2 tokens $\times$ 6 tones $\times$ 3 repetitions.

Table 2. Assimilation of Vietnamese tones into Mandarin tone categories by beginning and experienced learners.

Learner groups		Vietnamese stimulus					
		V44	V22	V35	V21	V415	V214
Response							
Beginning	M55	97 (5.6)*	26 (5.0)	–	8 (3.1)	–	–
	M35	–	–	90 (5.3)*	–	59 (3.7)	–
	M214	–	21 (4.3)	8 (3.4)	13 (3.5)	39 (3.1)	48 (4.8)
	M51	–	52 (5.2)*	–	75 (3.6)*	–	47 (4.7)
Assimilation		C	C	C	C	U _{-focalized}	U _{-clustered}
Experienced	M55	94 (5.9)*	24 (4.8)	–	12 (3.6)	–	–
	M35	–	–	93 (5.6)*	–	52 (3.9)	–
	M214	–	38 (4.6)	–	25 (4.2)	36 (3.8)	65 (5.1)*
	M51	–	35 (4.8)	–	60 (3.9)*	10 (3.4)	32 (4.8)
Assimilation		C	U _{-focalized}	C	C	U _{-focalized}	C

Notes. Categories in bold are choices that were selected significantly above chance (25%); ‘*’ = Categorized tone. Assimilations: C = Categorized, U = Uncategorized. Rating: 1 = poor, 7 = perfect; mean ratings are displayed in the parentheses. ‘–’ = no response or response < 5%.

2 Results

Table 2 shows the mean percent choice of categorizations to Mandarin tones and goodness of fit ratings for each Vietnamese tone. To determine PAM assimilation patterns, we followed two statistically-based criteria used in Chen et al. (2020) for Categorized assimilation, which are sensitive to variations among different native tone systems, instead of setting a fixed threshold (e.g. 50% or 70%). First, a native tone must be selected significantly above the chance level (i.e. 25%) as determined by *t*-tests. Second, that native category must be chosen significantly more often than any other response categories. To test this, we fitted the data using two linear mixed-effects models (one for each learner group) with percent choice as the dependent variable, native categories as a fixed factor, and participants as the random intercept.

To calculate the *p*-values for the fixed effects, we used the Kenward–Roger approximation to estimate the degrees of freedom, as recommended by Halekoh and Hojsgaard (2014), via the *Anova* function from the *car* package (Fox and Weisberg, 2019) in *R* (R Core Team, 2018), with test specified as ‘*F*’. We then ran multiple comparisons between native categories with the *R*-package *lsmeans* (Lenth, 2016). For statistical details, see Appendix A, Table A1–A3, and for individual variations in assimilation choices and ratings, see Appendix A, Table A9 in the online supplementary material in the html file at https://osf.io/jdey5/?view_only=cdacba08ec194015b2fcbbfd0ba27c99.

For the beginning learners, four Vietnamese tones were Categorized to Mandarin tone categories as we expected. V44 and V35 were Categorized as M55 and M35 respectively. V21 was also Categorized as M51. V22 was also Categorized as M51. Given that

both V22 and V21 were Categorized as M51, we further compared their goodness-of-fit ratings. The t -test indicated that there was a significant difference in goodness ratings, $t=3.63$, $df=38.99$, $p<0.001$, and thus the V22–V21 contrast was a Category–Goodness assimilation contrast.

For Uncategorized assimilations, we followed previous research (Chen et al., 2020; Faris et al., 2016, 2018) to distinguish three different ways in which an Uncategorized L2 phone can be perceived: (1) a *focalized* response in which the L2 phone is assimilated as primarily similar to a single L1 category but choices of that L1 category fall below the defined categorization threshold; (2) a *clustered* response in which the L2 phone is assimilated below the threshold to a small set of L1 categories; or (3) a *dispersed* response in which the assimilation of a non-native phone category is distributed across many L1 categories (all below the chance level). V415 assimilation was Uncategorized_{-focalized} because only M214 was selected above the chance level. V214 assimilation was Uncategorized_{clustered} and split between M51 and M214, both selected significantly above the chance level.

For the experienced group, the assimilation of V44, V35, V21 and V415 remained unchanged. The assimilation of V22 changed from being Categorized to being Uncategorized_{-focalized} as only M214 was selected significantly above the chance level. In addition, V214 changed from being Uncategorized to being Categorized as M214.

3 Discussion

Extending the PAM principle (Best, 1995) that a non-native/L2 phone can be an excellent to poor fit (gradient and phonetic aspect) to a native/L1 phonological category (categorical and phonological aspect), we propose that L1 *phonological* influence is generally stronger for Categorized than Uncategorized assimilations. In addition to phonological constraints, L2 listeners may retain some residual *phonetic* sensitivity to differences of the L2 phones from the L1 categories, as indicated separately by percent choice (in the assimilation task) and rating scores (in the goodness rating task). The Mandarin listeners showed strong evidence of Mandarin L1 phonological influences in assimilation of L2 Vietnamese tones along with varying degrees of residual phonetic sensitivity.

Three out of six Vietnamese tones were Categorized as native tone categories by both L2 learner groups. Two of them (V44 and V35) showed strong L1 phonological influences and weak residual phonetic sensitivity to differences between L1 and L2 tones. In addition, V21 was Categorized with lower percent choices and goodness ratings, indicating moderate residual phonetic sensitivity to the deviation of V21 from M51. This is in line with the ‘equivalence classification’ as in SLM (Flege, 1995). That is, some L2 features (pitch height in this case) that are not contrastive in the native language will be ‘ignored’ when categorizing a non-native phone to a native phoneme category. But listeners can still retain sensitivity to phonetic differences in other perception tasks.

V415, which is a falling–rising tone but differs from M214 phonetically, was the only Uncategorized Vietnamese tone that was perceived consistently by the two groups of learners. It was split between M35 and M214 with percent choice and goodness ratings in the medium range, suggesting moderate residual phonetic sensitivity to differences between V415 and both Mandarin categories.

Two Vietnamese tones were assimilated differently at the group level between beginning and experienced learners, indicating effects of L2 learning experience and further

perceptual attunement of the two L2 tones. V22, a low-level tone, was Categorized as M214 by beginning learners with lower percent choice compared with those for V44 and V35, suggesting moderate residual phonetic sensitivity to the phonetic differences between V22 and M214. This phonetic sensitivity could facilitate perceptual learning of the tone, which changed it to be an Uncategorized assimilation when perceived by some experienced learners. A new category could be established for this low-level tone, which is absent in the Mandarin tone system. However, there are also some individual differences (see Table A9 in the online supplementary material). In this case, some experienced learners consistently chose M51 or M55, but when averaged across participants, the mean percent for that category was low even though it was high for some individual participants.

On the other hand, V214 was Uncategorized by the beginning learners as it is phonetically similar to M214 and M51 in terms of the falling contour. Experienced learners stabilized the assimilation of V214 as M214 with moderate percent choice and high goodness ratings, suggesting relatively weak residual phonetic sensitivity to differences between V214 and M214. We speculate that experienced learners have modified their native M214 tone category to include V214, which could result in more stable and efficient perception of the L2 tone.

4 Predictions about discrimination

PAM (Best, 1995) stated that discrimination of a Two-Category contrast should be excellent and that of an Uncategorized–Categorized contrast should be very good because both involve perception across L1 phonological category boundaries. On the other hand, discrimination of a Category–Goodness contrast, which relies on within-category phonetic sensitivity with no assistance from an L1 phonological difference, may range from being moderate to good depending on the perceived phonetic difference of each L2 phone to the L1 category.

More recently, phonological overlap in L1 response categories between non-native contrasts has been integrated with assimilation types to fine-tune discrimination predictions (Best et al., 2019; Chen et al., 2023; Faris et al., 2016, 2018; Tyler et al., 2014). This is essential to account for discrimination variations of contrasts that belong to the same assimilation type (e.g. Two-Category, Uncategorized–Categorized, Category–Goodness, Uncategorized–Uncategorized). When two non-native phones are assimilated phonologically to the same set of L1 categories, i.e. *complete overlap*, discrimination will be poor due to strong L1 phonological interference. When a non-native contrast only overlaps on some of L1 categories, i.e. *partial overlap*, discrimination accuracy will increase but still be lower than when all members of a non-native contrast are assimilated to completely different sets of L1 categories, i.e. *non-overlap*, in which discrimination is facilitated by L1 phonological distinctions. In the above case, only overlaps of above-chance categories (which are bolded in Table 1) are considered, thus reflecting L1 *phonological* influences.

To make predictions for our study, we first identified assimilation types and phonological overlap types based on the perceptual assimilation results for each group. For the beginning learner group, V44–V22, V44–V35, V44–V21, V22–V35, and V35–V21

were Two-Category contrasts with no phonological overlap, i.e. TC-N, and discrimination of these contrasts would be excellent. V44–V415, V44–V214, V22–V415, V35–V214, and V21–V415 were Uncategorized–Categorized contrasts with no phonological overlap, i.e. UC-N (N stands for no overlap), and discrimination of these contrasts would be good. V415–V214 was an Uncategorized–Uncategorized contrast with no phonological overlap, i.e. UU-N, and its discrimination should also be good. Given that TC-N, UC-N, and UU-N contrasts involve perception across a native phonological category boundary, discrimination of these contrasts should be generally comparable.

V22–V214, V35–V415, and V21–V214, were Uncategorized–Categorized contrasts but with partial phonological overlap, i.e. UC-P (P stands for partial overlap), their discrimination should be poorer than TC-N, UC-N, and UU-N. V22–V21 was a Category–Goodness contrast, and its discrimination should also be poorer than TC-N, UC-N, and UU-N contrasts, $\{TC-N = UC-N = UU-N\} > \{TC-P = UC-P = UU-P = CG\}$.

For the experienced learner group, the assimilation and phonological overlap types of some contrasts involving V22 and V214 changed. Our predictions of V44–V22, V44–V214, V22–V35, and V35–V214 remain the same as they only involve changes between TC-N and UC-N assimilation. Discrimination of V22–V415 (UC-N $\rightarrow$ UU-N) should also be comparable between the beginning and experienced learner groups. On the other hand, the predicted accuracy for discrimination will change for V21–V214 and V22–V21. Discrimination of both V21–V214 (UC-P $\rightarrow$ TC-N) and V22–V21 (CG $\rightarrow$ UC-N) should be improved due to less phonological overlap. Unlike most previous studies, which selected a few typical contrasts to test, we tested all 15 tone contrasts of the six Vietnamese tones, forming 105 possible pairs for comparison. (For the full list of comparisons and relevant predictions, see Appendix A, Table A4 in the online supplementary material.)

In addition to phonological overlap types, listeners' phonetic sensitivity to contrasts can vary as a function of perceived *phonetic* similarities of the contrasting tones, as indicated by percent choice and goodness ratings. One approach for considering residual phonetic sensitivity is by calculating *overlap scores* (as in Flege and MacKay, 2004; Levy, 2009), which is the summed percentages of all overlapping native category choices, including *below-chance* choices, for a non-native contrast. Goodness ratings, however, were not considered in the above overlap score formulae. Given that both goodness ratings and percent choice reflect residual phonetic sensitivity (Chen et al., 2020), we also multiplied percent choice by goodness ratings to calculate a *fit-index* (Wu et al., 2014) and the *fit-index difference score*, i.e. a difference score of fit-index between two tones of an L2 contrast. Note that a high *fit-index difference score* indicates *high* residual phonetic sensitivity to differences between two tones whereas a high *overlap score* shows *low* listener sensitivity (for both scores for each contrast for both learner groups, see Appendix A, Table A5–A8 in the online supplementary material).

To determine whether there were any significant differences of each score between each pair of contrasts, we fitted the overlap scores and fit-index difference scores as the dependent variables into separate Linear Mixed Effect Regression (LMER) models (for each measure of phonetic overlap) with contrasts as fixed factors, and participants as the random intercept. Thus, we built two separate models for each learner group. Then, we ran pairwise multiple comparisons with Tukey adjustments for each tone pair contrast

(for statistical details and relevant predictions, see Appendix A, Table A5–A8 in the online supplementary material). If the difference of phonetic overlap indices between two contrasts of a pair is significant, we predict that the two contrasts should be discriminated differently. Note that a high *fit-index difference score* predicts *high* discrimination accuracy whereas a high *overlap score* predicts *low* discrimination accuracy.

III Experiment 2

To test the above predictions, we ran AX discrimination tests for each learner group. The same participants were tested online as in Experiment 1 and were asked to take the test in a quiet environment with the same headphones they used for Experiment 1. The intensity of all auditory stimuli was normalized as in Experiment 1.

I Method

An AX discrimination task ('same-different') was used as in many previous discrimination studies (Asano, 2018; Werker and Logan, 1985; Werker and Tees, 1984). On each trial, participants heard stimulus A followed by a silence of 1,000 ms inter-stimulus interval (ISI). After the ISI, the second stimulus X was presented and the prompt 'Same or different?' appeared on the screen, signaling the participant to respond as quickly as possible. Half of the trials consisted of 'same tone' pairs, while the other half were 'different tone' pairs, presented in random order. For the 'same tone' trials, the two stimuli were spoken by the same speakers and were physically identical.

Before the discrimination test, participants completed 20 practice trials with the same syllables produced by a third Vietnamese speaker which were not used in the test session. In the test session, each participant completed 480 trials in total, i.e. 2 informants $\times$ 2 syllables (/ma/ and /mi/) $\times$ 2 tokens $\times$ 15 contrasts $\times$ 4 AB conditions. No feedback was given in either the practice or test session.

2 Results

In order to minimize response bias effects on the AX data, we calculated d-prime or d' (Macmillan and Creelman, 2005) for discrimination of each tone pair (see Figure 2). Adjustments were made for probabilities of 0 and 1 to avoid infinite values by adding 0.5 to all raw data cells, regardless of whether zeroes are present (Hautus, 1995; Miller, 1996). A 'hit' is defined as the number of correct 'different' responses on AB or BA trials, while a 'false positive' is defined as the number of incorrect 'different' responses on AA or BB trials.

We fitted the data using a linear mixed-effects model with d' as the dependent variable, and learning experience (beginning versus experienced learners), and tone contrasts ('V44–V22', 'V44–V35', 'V44–V21', 'V44–V415', 'V44–V214', 'V22–V35', 'V22–V21', 'V22–V415', 'V22–V214', 'V35–V21', 'V35–V415', 'V35–V214', 'V21–V415', 'V21–V214', 'V415–V214') as fixed factors, and participants as a random intercept (for the coefficients of the fixed effects and interactions, see Table B1 in the Appendix B in the online supplementary material).

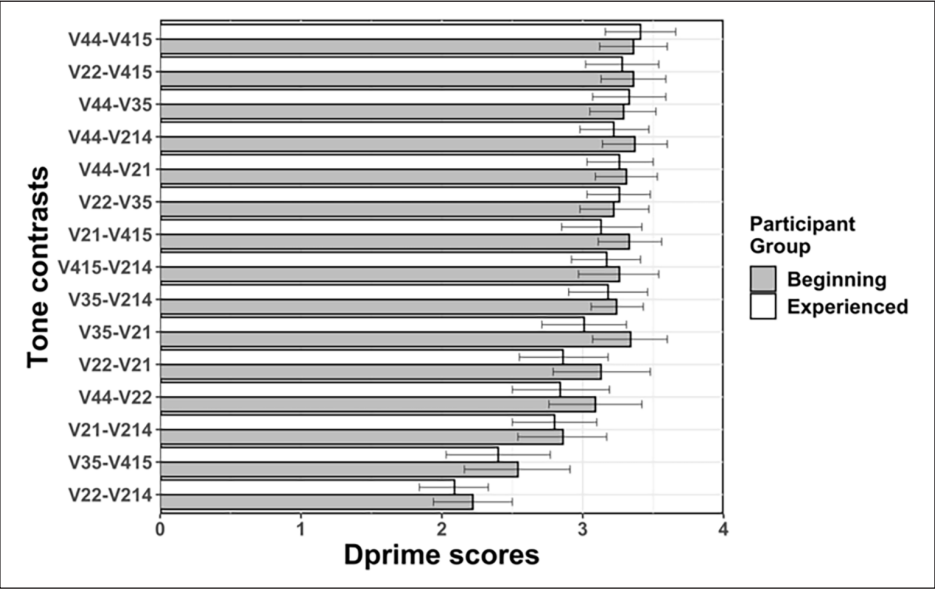


Figure 2. Discrimination of Vietnamese tone contrasts by Mandarin learners of Vietnamese.

To calculate the *p*-Values for the fixed effects, we again used the Kenward–Roger degrees of freedom approximation (Halekoh and Hojsgaard, 2014), and the *Anova* function from the *car* package in *R*, with test specified as ‘F’. Note that what we report here are main effects for multilevel factors or interactions, rather than comparisons with the baseline level using *t*-tests. Thus, the reported main and interaction effects were averaged across all levels of the other factors, allowing them to be directly utilized for testing our predictions. There was a significant main effect of tone contrasts, $F(14, 658)=28.4$, $p<.001$, but no main effect of learning experience. The learning experience by tone contrast interaction was also non-significant.

To test our separate predictions for the two participant groups, we built two separate models (one for each participant group) with *d'* as the dependent variable, and tone contrasts as fixed factors, and participants as a random intercept. Then, we ran pairwise multiple comparisons with Tukey adjustments using the *lsmeans* package as in our previous analysis for the 105 tone contrast pairs. For a given pair of contrasts, if their discrimination accuracies (*d'*) are not significantly different, we consider their discrimination to be comparable. Otherwise, one contrast of the pair is discriminated better or poorer than the other of the pair (for the full list of results and statistical details, see Appendix B, Tables B2 and B3 in the supplementary material).

We compared the discrimination results with reference to our three sets of predictions (see Tables 3 and 4 for general patterns, and see Appendix B, Tables B4 and B5 for details about accuracies for each contrast type). For the beginning learner group, among the contrasts that were *phonologically* predicted to be equally easy to distinguish, only

Table 3. Phonological (based on assimilation and overlap types) and phonetic (based on overlap scores or fit-index difference scores) predictions of discrimination of Vietnamese contrasts by Mandarin beginning learners and test results.

Predictions			Results	Cases
Phonological	Overlap	Fit-index		
=	=	=	=	31
=	=	>	=	19
=	>	>	=	3
=	<	=	=	3
=	<	<	=	2
=	=	=	>	2
=	=	=	<	1
>	>	>	>	21
>	>	>	=	10
<	<	<	<	7
<	<	<	=	6

Notes. =: equally good discrimination; >: better discrimination of contrast A than contrast B; <: poorer discrimination of contrast A than contrast B.

Table 4. Phonological (based on assimilation and overlap types) and phonetic (based on overlap scores or fit-index difference scores) predictions of discrimination of Vietnamese contrasts by Mandarin experienced learners and test results.

Predictions			Results	Cases
Phonological	Overlap	Fit-index		
=	=	=	=	29
=	=	>	=	18
=	=	<	=	2
=	>	>	=	11
=	<	=	=	1
=	<	<	=	9
=	>	>	>	7
=	<	=	<	1
=	<	<	<	1
>	>	>	>	15
>	=	=	>	2
<	<	<	<	5
<	<	=	<	2
<	=	=	<	1
<	=	=	=	1

Notes. =: equally good discrimination; >: better discrimination of contrast A than contrast B; <: poorer discrimination of contrast A than contrast B.

three cases did not align with these predictions (see Appendix B, Table B6 in the online supplementary material). These discrepancies occurred mostly in comparisons between UC-P (V22–V214; V35–V415) and CG (V22–V21) contrasts, as well as within the UC-P contrasts, between V22–V214 and V21–V214. On the other hand, for the contrasts that were *phonologically* predicted to exhibit either better or worse discrimination, there were 16 cases that did not conform to these predictions. These instances mainly encompassed comparisons between UC-P (V21–V214)/CG (V22–V21) and TC-N, UC-N, UU-N contrasts.

For the experienced learner group, among the contrasts that were *phonologically* predicted to be equally easy to discriminate, nine cases deviated from these predictions (see Appendix B, Table B7 in the online supplementary material). However, these discrepancies were successfully predicted by either overlap or fit-index difference scores. On the other hand, for the contrasts that were *phonologically* predicted to exhibit different discrimination accuracies, there was only one case that did not conform to these predictions, and that case was successfully predicted by both overlap and fit-index difference scores.

In general, the overall accuracies (percentages of correct predictions) were higher for the phonological predictions based on assimilation and overlap types (beginning learners=82%, experienced learners=90%) relative to pure phonetic predictions based on either overlap scores (beginning learners=74%, experienced learners=77%) and fit-index difference scores (beginning learners=59%, experienced learners=56%). However, it should be noted that these findings are descriptive with the aim to summarize the general patterns and have not been statistically tested for significance.

3 Discussion

In this experiment, our phonological predictions based on assimilation types and phonological overlap types were generally upheld for both participant groups and outperformed two sets of predictions based on pure phonetic overlap measures, suggesting the importance of L1 phonological influences on L2 tone perception. In addition, L2 listeners did show some residual phonetic sensitivities that modulated L1 phonological influences.

For the beginning learner group, we successfully predicted the TC-N×TC-N (‘×’=a comparison between two contrasts), TC-N×UC-N, TC-N×UU-N, UC-N×UC-N, UC-N×UU-N to be discriminated with comparable accuracy (see Appendix B, Table B6 in the online supplementary material). The majority of the contrast pairs that deviated from our predictions involved the CG contrast V22–V21. It was predicted to be discriminated worse than TC-N, UC-N, and UU-N. However, V22–V21 was discriminated as good as TC-N, UC-N and UU-N. V21 received relatively low goodness ratings with a high percent choice. The discrepancy between percent choice and goodness ratings suggests that listeners attended to the phonetic differences of the non-native tone from their native categories, although they categorized V21 as M51. This phonetic sensitivity could increase their discrimination performance to be even better than some UC-P contrasts, such as V22–V214, and V35–V415. In addition, the AX discrimination task taps more into a phonetic level distinction relative to other tasks, such as AXB. The overall performance is good among beginning learners, which makes some TC-N, UC-N, and UU-N contrasts less distinguishable from the CG and UC-P contrasts.

For the experienced learner group, most of the cases that did not align with our predictions based on the assimilation and phonological overlap types involved the TC-N, UC-N and UU-N contrasts (see Appendix B, Table B7 in the online supplementary material). Instead of being discriminated equally well, their discrimination accuracies were different but all predicted by the overlap scores. The above findings suggest that L1 phonological influences as indicated in assimilation and overlap types played an important role in L2 discrimination of tone contrasts while L2 listeners did retain residual phonetic sensitivities to L1 and L2 tone categories and these phonetic influences also affected L2 tone perception.

We did not find a main effect of L2 learning experience on discrimination accuracy nor any interactions between learning experience and tone contrasts. This is similar to a previous study with English native learners perceiving Mandarin tone contrasts (Hao, 2017). In our case, both groups displayed relatively good discrimination performance for contrasts that involved perception across L1 phonological category boundaries as we expected. For those challenging L2 contrasts with less assistance from L1 phonological effects, such as V35–V415, the experienced learner group did not demonstrate superior performance compared to the beginning learner group. This might be attributed to the experienced learners' limited exposure to Vietnamese in authentic communicative contexts.

IV General discussion

The aims of the two experiments in the present study were to examine L1 phonological and phonetic influences on the perceptual assimilation and discrimination of non-native lexical tones and to determine whether and how L2 learning experience modulates these L1 effects.

I Native language phonological and phonetic effects

First of all, the L1 tone system constrained the perceptual assimilation of L2 tones for both learner groups. Both beginning and experienced learners Categorized V44, V35, V21 into L1 tone categories, suggesting strong L1 phonological influences in perceptual assimilation. Both V44 and V35 were Categorized with high percent choice and goodness ratings, indicating weak residual phonetic sensitivity to differences between the two tones and native tones. Thus, according to PAM, no perceptual learning will occur for the two tones.

Despite strong L1 phonological effects, both learner groups retained residual phonetic sensitivity to differences between L1 and L2 tone categories. The low falling tone V21 was Categorized as the only Mandarin falling tone M51 with medium percent choice and goodness ratings, suggesting moderate residual phonetic sensitivities, which should facilitate the development of a new category for this tone in the common L1-L2 sound system.

Second, our phonological predictions about discrimination accuracy based on assimilation and phonological overlap types were largely upheld and generally outperformed the two pure phonetic overlap measures for both learner groups. The reason may be that phonetic overlap measures, particularly the fit-index difference scores, rely too much on phonetic details, and thus lack the power of capturing phonological level distinctions.

Given that not all the perceived phonetic differences are phonologically contrastive, our participants, as native tone language speakers, maintain L2 phonological contrasts by ignoring some phonetic differences. On the other hand, within the same assimilation-overlap category, overlap scores and/or fit index difference scores successfully fine-tuned some predictions. For the experienced group, overlap scores successfully predicted some TC-N $\times$ UC-N contrasts to be discriminated with different accuracies. Thus, L1 influences exist at both phonological and phonetic levels and need to be considered thoroughly.

2 L2 learning experience effects

Perceptual assimilation of two Vietnamese tones, V22 and V214, was different across the two groups. Categorized L2 phones as perceived by beginning learners could turn to be Uncategorized phones as these learners gain more experience with L2, provided that these learners could initially perceive or gradually detect at least some phonetic differences between the L2 phone and the corresponding L1 phoneme. V22, a low-level tone, was initially Categorized by beginning learners as M214, which can be phonetically realized as a low falling tone. The percent choice, however, was not very high, suggesting overall moderate residual phonetic sensitivity to the differences between V22 and M214. When experienced learners perceived this tone, they assimilated it as Uncategorized, split between M214 and M51 with low percent choice, suggesting low native phonological constraints and high residual phonetic sensitivity. This may indicate that experienced learners detected increased phonetic distances between L1 and L2 categories. We infer that at least some experienced learners may have developed a new phonological category for this tone, which should contrast with the high-level tone, V44. Since Mandarin does not have a phonological contrast for level tones, V22 cannot be assimilated phonologically as any single Mandarin native category but assimilated phonetically as similar to M214 and M51. This is similar to the dissimilation process in SLM/SLM-r, in which new L2 categories shift away from neighboring L1 categories to maintain phonetic contrast between pairs of L1 and L2 categories (Flege and Bohn, 2021).

On the other hand, an Uncategorized L2 phone that is assimilated to a set of L1 categories initially can be Categorized into one of them later, if listeners develop a common L1-L2 category from that L1 category. V214 was initially Uncategorized by beginning learners. With more exposure to the language, experienced learners Categorized V214 as M214 with relatively high goodness ratings. We speculate that experienced learners have established a shared category in the common L1-L2 space by perceptually modifying their native M214 tone category to incorporate V214. This perceptual development is also referred to as an assimilation process by SLM/SLM-r, in which L2 learners develop a merger of the phonetic properties of an L1 category and the L2 category to which it remains perceptually linked (Flege and Bohn, 2021).

However, for tone discrimination, no overall significant differences were found between the two learner groups nor any interactions involving learner groups as suggested by the linear mixed-effects model. The fact that the experienced learners did not outperform the beginning learners in discriminating L2 contrasts is in line with some

previous L2 tone perception studies (Hao, 2017). For the present study, there are some factors that can account for this lack of significant effects. First, although AX discrimination is a good task to test L1 influences on L2 tone discrimination as different tone contrasts did show different discrimination accuracies within each group, it may not be sensitive enough to reveal between-group differences, particularly for contrasts that involve L2 tones across L1 phonological boundaries, such as TC and UC with no overlapping L1 response categories. More than half of the contrasts that we tested belonged to this type. In these cases, both groups can benefit from their L1 phonological and phonetic effects and thus their performance for these tones was not significantly different.

On the other hand, for contrasts with partial overlapping in L1 response categories, i.e. potentially difficult contrasts, such as V35–V415, the experienced group did not outperform the beginning group. Note that our experienced learner group only had two years of formal classroom-based learning experience (explicit learning) rather than natural exposure to Vietnamese in communicative settings (implicit learning). L2 speech learning models, such as SLM-r, argue that both quality and quantity of L2 input are crucial for the formation of language-specific L2 phonetic categories. The SLM-r specifically defines ‘phonetic input as the sensory stimulation associated with L2 speech sounds that are heard and seen during the production by others of L2 utterances in meaningful conversations’ (Flege and Bohn, 2021). It has been found that immersion experience with L2 could significantly improve discrimination of L2 contrasts (Jia et al., 2006). In addition, Japanese participants with intensive English conversation experience showed identification and discrimination patterns of American English approximants that were more similar to native performance than those who had had little English conversation experience (Best and Strange, 1992). Thus, the limited availability of high-quality natural input might explain the slow perceptual development of the challenging contrasts observed in our study. It would be interesting to compare students with and without study-abroad experience in terms of their perception of these difficult tone contrasts.

3 Limitations and future directions

First, we did not include a naive listener group to establish a true baseline for early learning. While we have observed cross-sectional development patterns, future research could include a group of naive listeners to better understand initial learning of tone categories.

Second, to fully understand second language tone development, future research should investigate the production of Vietnamese tones by Mandarin learners with differing L2-Vietnamese experiences. The relationship between perception and production in L2 development is notably complex (Baese-Berk, 2019; Kirby and Giang, 2021; Laméris and Post, 2023). Some production studies did detect effects of L2 experience on L2 speech learning (Turner, 2023). A study on L2 perception and production of voiceless stops by English learners of Spanish (Gorba and Cebrian, 2021) found that the L2 speakers were better at producing than perceiving Spanish stops, with L2 experience having a positive effect on production but no effect on perception. Additional L2 input may further attune some aspects of L2 tone categories, which may be more clearly reflected in learners’ production improvement rather than that of discrimination. Some of the Vietnamese tones, on the other hand, can be easily perceived but remain difficult to produce.

V Conclusions

In summary, the current findings indicate that L1 phonological and phonetic factors affect L2 perception of lexical tones. PAM-driven predictions based on assimilation and phonological overlap types generally outperformed predictions based purely on phonetic overlap measures when accounting for variations in discrimination performances. On the other hand, phonetic overlap scores are useful in fine-tuning predictions of contrasts with the same assimilation and phonological overlap type. L2 learning experience affected perceptual assimilation patterns but not discrimination. Future studies could further examine how differing L2 learning experiences, such as learning in a classroom setting versus in a native environment, affect L2 tone perception and production.

Acknowledgements

We would like to thank Professor Boping Yuan, Dr Lilong Xu, and three anonymous reviewers for their invaluable comments on earlier drafts of this article.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this work was supported by a China Foreign Language Education project (grant number ZGWYJYJJ11Z007).

ORCID iD

Juqiang Chen  <https://orcid.org/0000-0001-8151-3964>

Supplemental material

Supplemental material for this article is available online.

Notes

1. ‘Categorized’ with a capital ‘C’ is used intentionally to denote that the assimilation meets PAM’s standards for Categorization.
2. The designation of high versus low features in this contrast is based on their phonological behavior in reduplication (Yip, 2002). If the inputs are rising or high falling–rising tone (*hôi*), the prefixal reduplicant surfaces as the high-level tone, whereas if the inputs are falling or low falling–rising (*ngã*), then the prefixal reduplicant surfaces as low-level tone. The phonological features and Chao numbers are thus not consistent in this case.
3. In future research, more homogeneous participants would be desirable, e.g. recruiting only Beijing Mandarin speakers. For practical reasons, however, strict control of dialect backgrounds of L2 Vietnamese learners in China was not possible because Vietnamese programs are offered only at a few Chinese universities and recruit a limited number of students each year. With this said, evidence on the effect of dialect backgrounds on L2 perception is mixed (Liu and Escudero, 2023; Zhang et al., 2022). In the present study, we used participants’

assimilation patterns to directly predict and account for variations in their discrimination accuracy. In this way, potential dialect effects should be consistent across assimilation and discrimination tasks.

References

- Antoniou M, Best CT, and Tyler MD (2013) Focusing the lens of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *The Journal of the Acoustical Society of America* 133: 2397–411.
- Asano Y (2018) Discriminating non-native segmental length contrasts under increased task demands. *Language and Speech* 61: 409–29.
- Baese-Berk MM (2019) Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics* 81: 981–1005.
- Best C (1995) A direct realist view of cross-language speech perception. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 171–204.
- Best C (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: Article 364.
- Best C, Avesani C, Tyler MD, and Vayra M (2019) PAM revisits the articulatory organ hypothesis: Italians' perception of English anterior and Nuu-Chah-Nulth posterior voiceless fricatives. In: Nyvad AM, Hejná M, Højen A, Jespersen AB, and Sørensen MH (eds) *A sound approach to language matters: In honor of Ocke-Schwen Bohn*. Aarhus: Aarhus University, pp. 13–40.
- Best C, McRoberts GW, and Goodell E (2001) Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America* 109: 775–94.
- Best C and Strange W (1992) Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20: 305–30.
- Best CT and Tyler MD (2007) Nonnative and second-language speech perception. In: Bohn OS and Munro MJ (eds) *Language experience in second language speech learning*. Amsterdam: John Benjamins, pp. 13–34.
- Boersma P (2001) Praat, a system for doing phonetics by computer. *Glott International* 5: 341–45.
- Braun B, Galts T, and Kabak B (2014) Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. *Second Language Research* 30: 323–50.
- Braun B and Johnson EK (2011) Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics* 39: 585–94.
- Browman CP and Goldstein L (1989) Articulatory gestures as phonological units. *Phonology* 6: 201–51.
- Brunelle M (2009) Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37: 79–96.
- Bundgaard-Nielsen RL, Best C, Kroos C, and Tyler MD (2012) Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics* 33: 643–64.
- Bundgaard-Nielsen RL, Best C, and Tyler MD (2011a) Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition* 33: 433–61.
- Bundgaard-Nielsen RL, Best C, and Tyler MD (2011b) Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics* 32: 51–67.

- Burnham D, Kasisopa B, Reid A, et al. (2015) Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics* 36: 1459–91.
- Chan RK and Leung JH (2020) Why are lexical tones difficult to learn? *Studies in Second Language Acquisition* 42: 33–59.
- Chang YS, Yao Y, and Huang BH (2017) Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America* 141: EL120–26.
- Chao YR (1930) A system of tone-letters. *Le Maître Phonétique* 45: 24–27.
- Chao YR (1968) *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.
- Chen J, Antoniou M, and Best CT (2023) Phonological and phonetic contributions to perception of non-native lexical tones by tone language listeners: Effects of memory load and stimulus variability. *Journal of Phonetics* 96: 101199.
- Chen J, Best CT, and Antoniou M (2020) Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners. *Journal of Phonetics* 83: 101013.
- Escudero P (2005) *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Ridderkerk: LOT.
- Faris MM, Best C, and Tyler MD (2016) An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America* 139: EL1–5.
- Faris MM, Best C, and Tyler MD (2018) Discrimination of uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics* 70: 1–19.
- Flege JE (1995) Second-language speech learning: Theory, findings, and problems. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 229–73.
- Flege JE and Bohn O-S (2021) The revised speech learning model (SLM-r). In: Wayland R (ed.) *Second language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press, pp. 3–83.
- Flege JE and MacKay IRA (2004) Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26: 1–34.
- Fox J and Weisberg S (2019) *An R companion to applied regression* (3rd ed.). Sage. Available at: <https://www.john-fox.ca/Companion> (accessed April 2025).
- Gorba C and Cebrian J (2021) The role of L2 experience in L1 and L2 perception and production of voiceless stops by English learners of Spanish. *Journal of Phonetics* 88: 101094.
- Halekoh U and Hojsgaard S (2014) A Kenward–Roger approximation and parametric bootstrap methods for tests in linear mixed models: The R package pbrtest. *Journal of Statistical Software* 59: 1–30.
- Hao Y-C (2012) Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics* 40: 269–79.
- Hao Y-C (2017) Second language perception of Mandarin vowels and tones. *Language and Speech* 61: 135–52.
- Hautus MJ (1995) Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, and Computers* 27: 46–51.
- Jia G, Strange W, Wu Y, Collado J, and Guan Q (2006) Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America* 119: 1118–30.
- Kirby J and Giang ĐL (2021) Relating production and Perception of L2 Tone. In: Wayland R (ed.) *Second language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press, pp. 249–72.

- Kuhl PK and Iverson P (1995) Linguistic experience and the ‘perceptual magnet effect’. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 121–54.
- Laméris TJ and Post B (2023) The combined effects of L1-specific and extralinguistic factors on individual performance in a tone categorization and word identification task by English-L1 and Mandarin-L1 speakers. *Second Language Research* 39: 833–71.
- Lenth RV (2016) Least-squares means: The R package lsmeans. *Journal of Statistical Software* 69: 1–33.
- Levy ES (2009) On the assimilation-discrimination relationship in American English adults’ French vowel learning. *The Journal of the Acoustical Society of America* 126: 2670–82.
- Liu X and Escudero P (2023) How bidialectalism affects non-native speech acquisition: Evidence from Shanghai and Mandarin Chinese. *Applied Psycholinguistics* 44: 969–90.
- Lobanov BM (1971) Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49: 606–08.
- Macmillan NA and Creelman CD (2005) *Detection theory: A user’s guide*. New York: Psychology Press.
- Mennen I (2015) Beyond segments: Towards a L2 intonation learning theory. In: Delais-Roussarie E, Avanzi M, and Herment S (eds) *Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations*. Berlin: Springer, pp. 171–88.
- Miller J (1996) The sampling distribution of d’. *Perception and Psychophysics* 58: 65–72.
- Nhàn NT (1984) The syllabeme and patterns of word formation in Vietnamese. Unpublished PhD dissertation, New York University, New York, NY, USA.
- Peirce J, Gray JR, Simpson S, et al. (2019) PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51: 195–203.
- Pelz E, Liu J, and Qi C (2022) Native language experience with tones influences both phonetic and lexical processes when acquiring a second tonal language. *Journal of Phonetics* 95: 101197.
- Qin Z, Tremblay A, and Zhang J (2019) Influence of within-category tonal information in the recognition of Mandarin-Chinese words by native and non-native listeners: An eye-tracking study. *Journal of Phonetics* 73: 144–57.
- R Core Team (2018). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org> (accessed April 2025).
- Reid A, Burnham D, Kasisopa B, et al. (2015) Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention Perception and Psychophysics* 77: 571–91.
- So CK and Best C (2011) Categorizing mandarin tones into listeners’ native prosodic categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics* 47: 133–45.
- So CK and Best C (2014) Phonetic influences on English and French listeners’ assimilation of mandarin tones to native prosodic categories. *Studies in Second Language Acquisition* 36: 195–221.
- Tsukada K and Han J-I (2019) The perception of Mandarin lexical tones by native Korean speakers differing in their experience with Mandarin. *Second Language Research* 35: 305–18.
- Tsukada K, Kondo M, and Sunaoka K (2016) The perception of Mandarin lexical tones by native Japanese adult listeners with and without Mandarin learning experience. *Journal of Second Language Pronunciation* 2: 225–52.
- Turner J (2023) Phonetic development of an L2 vowel system and tandem drift in the L1: A residence abroad and L1 re-immersion study. *Language and Speech* 66: 756–85.
- Tyler MD, Best C, Faber A, and Levitt AG (2014) Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71: 4–21.
- Werker JF and Logan JS (1985) Cross-language evidence for three factors in speech perception. *Perception and Psychophysics* 37: 35–44.

- Werker JF and Tees RC (1984) Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America* 75: 1866–78.
- Wong PCM and Perrachione TK (2007) Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics* 28: 565–85.
- Wu X, Munro MJ, and Wang Y (2014) Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics* 46: 86–100.
- Yip M (2002) *Tone*. Cambridge: Cambridge University Press.
- Yu K, Li L, Chen Y, et al. (2019) Effects of native language experience on Mandarin lexical tone processing in proficient second language learners. *Psychophysiology* 56: Article e13448.
- Zhang Y, Zheng Y, and Li G (2022) Dialect effects on Mandarin tone perception development. *Language and Speech* 65: 681–96.