Laboratory Phonology
Journal of the Association for Laboratory Phonology

Chen, J., Best, C. T., & Antoniou, M. (2023). Phonological and phonetic contributions to Thai-naïve Mandarin and Vietnamese speakers' imitation of Thai lexical tones: Effects of memory load and stimulus variability. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 14(1), pp. 1–41. DOI: https://doi.org/10.16995/labphon.6435

Open Library of Humanities

Phonological and phonetic contributions to Thai-naïve Mandarin and Vietnamese speakers' imitation of Thai lexical tones: Effects of memory load and stimulus variability

Juqiang Chen*, Shanghai Jiao Tong University, School of Foreign Languages, China, juqiang.c@sjtu.edu.cn

Catherine T. Best, Western Sydney University, The MARCS Institute for Brain Behaviour and Development, AU; Haskins Laboratories, New Haven CT, USA, C.Best@westernsydney.edu.au

Mark Antoniou, Western Sydney University, The MARCS Institute for Brain Behaviour and Development, AU, m.antoniou@westernsydney.edu.au

*****Corresponding author.

The present study examined how native phonological and phonetic factors in non-native speech perception (Perceptual Assimilation Model [PAM]: Best, 1995) affect non-native imitation of Thai tones by Thai-naïve Mandarin and Vietnamese participants, and how memory load and stimulus variability shift the balance between phonological versus phonetic modes (Automatic Selective Perception model [ASP]: Strange, 2011) in imitation. Although overall imitation was quite good, native phonological influences, as reflected in Categorised versus Uncategorised assimilation types in a prior perception study with the same participants (Chen, Antoniou, & Best, 2023), constrained non-native tone imitations. Residual perceptual sensitivity to phonetic differences between the target stimuli and native tones, as reflected in percent choice and goodness ratings in assimilation, also affected imitation. Effects of stimulus variability and memory load were restricted and interacted with specific tones for each participant group. Imitations were generally more accurate under low memory load, constant talker and vowel conditions, consistent with a more phonetic mode of perception. Imitation accuracy decreased under high memory load, variable talker and vowel conditions, consistent with a more phonological mode of perception. Results support the PAM and ASP accounts of native phonological and phonetic effects in non-native perception and extend them to non-native tone imitation.

1. Introduction

To learn to speak a second language (L2), language learners not only need to perceive L2 phonemes but also to produce them. In L2 immersion contexts and instruction with native speakers and/or native speech materials, naïve listeners or early L2 learners start producing L2 words by imitating native speaker productions; that is, they reproduce the specific phonetic details of target items as closely as possible (Carignan, 2018). The materials that naïve listeners imitate in these cases are words in a language they do not know. This differs from repetition, which involves native speakers (or possibly proficient L2 speakers) reproducing the more abstract phonological identity of known target words/phrases (Nguyen, Dufour, & Brunellière, 2012).

Non-native imitation relies on an implicit relationship between non-native speech perception and production. Without accurate perceptual "targets", non-native production will be inaccurate, as posited by the Speech Learning Model (SLM, Flege, 1995). Thus, imitation provides an excellent basis for examining how perception influences production in non-native speech learning, while avoiding the confounding effects of orthographic knowledge as when printed targets are presented.

1.1. Native language constraints on speech imitation

How non-native perception influences production, especially in beginning L2 learners, is a central theoretical issue, but is as yet unresolved. Several L2 speech learning theories have made explicit or implicit claims about this. SLM (Flege, 1995) claims that the native language's influence on non-native production can be traced back partially to how a non-native phone is perceived. If it is equivalence classified as either identical or similar to the acoustically closest native (L1) category (category assimilation), a single phonetic category will be used to process the perceptually linked L1 and L2 phones, referred to as diaphones. In this case, a "merged" category will develop over time that subsumes the phonetic properties of both diaphones (Flege, Schirru, & MacKay, 2003). On the other hand, if a non-native phoneme is new to the native phonological system, it will be substituted by a range of variants in the early stages of L2 learning. Ultimately, however, a separate L2 phonetic category will be established. In this case, SLM predicts that the newly established L2 category and the nearest L1 speech category will deflect from each other in the learner's phonetic space (category dissimilation).

The Perceptual Assimilation Model (PAM; Best, 1995) was originally created to account for native language influences in non-native speech perception by naïve listeners. The more recently developed PAM-L2 (Best & Tyler, 2007) extends PAM principles to account for L2 speech learning, as well as implicitly extending them to production due to its direct-realist theoretical assumption that perceivers detect articulatory information in speech (Best, 1995; Fowler, 1986). PAM considers native language influences on non-native perception at both phonological and phonetic levels. If a given non-native phone is perceived as corresponding to a

single native phonological category, it is considered a Categorised assimilation. Yet within that native phonological constraint, listeners will nevertheless display residual sensitivity to within-category phonetic variations from the native phoneme it is assimilated to, commensurate with the magnitude of phonetic discrepancy from "good" native exemplars.

In Categorised assimilation, if the non-native phone is perceived as a good exemplar of that native category, then no further perceptual learning will occur for that non-native phone during L2 acquisition according to PAM-L2. But if it is perceived as a phonetically deviant exemplar of that native category, further learning will be possible. On the other hand, if a non-native phone is not assimilated cleanly to any single native phonological category (UnCategorised assimilation), it will be less susceptible to native language influence and be easier to learn in an L2, because UnCategorised assimilations reflect weaker native phonological influence than do Categorised assimilations.

In previous studies of native speakers, native phonological categories were reported to constrain reproduction of items along synthetic consonant and vowel continua. Native speakers failed to reproduce within-category phonetic variations among a continuum's stimulus items. Instead, their reproductions of the stimuli reflected their native phonological categories. For example, when English and Spanish monolinguals reproduced a synthetic stop consonant voice onset time (VOT) continuum ranging from /da/ to /ta/, neither group's productions showed a linear incremental increase in VOT, as the stimuli did. Instead, they reflected the VOT categories observed in their perception of their different native stop voicing category boundaries (Flege & Eefting, 1988). Similarly, when Finnish children and adults reproduced the items along a synthetic Finnish /æ/ to /ɑ/ vowel continuum, they showed categorical production patterns that matched their perception of native phonological categories along the continuum (Alivuotila, Hakokari, Savela, Happonen, & Aaltonen, 2007).

Imitation of non-native phonemes is also modulated by participants' native phonological systems. For example, when asked to imitate eight American English vowels /i, ɪ, e, ɛ, æ, ʌ, ɑ, u/, native Mandarin English L2 speakers showed the influence of their native language (Jia, Strange, Wu, Collado, & Guan, 2006): /ɛ/ and /æ/ have no Mandarin counterparts and were imitated less accurately than /i/ and /u/, which exist in Mandarin. Moreover, they displayed a positive correlation between perception and production accuracy, suggesting that non-native phones that are difficult to relate perceptually to L1 phonemes are also difficult to imitate.

Although there is ample evidence that native phonology constrains non-native imitation, there is evidence suggesting participants can bypass native phonological influence and produce phonetically accurate reproductions, at least in their native language. When presented with VOT-extended English voiceless stops as stimuli, native English speakers accurately reproduced them also with lengthened VOTs, compared to their baseline production of these consonants (Shockley, Sabadini, & Folwer, 2004).

Most repetition and/or imitation studies have examined perception-production relationships for consonants and vowels. Although lexical tones exist in about 70% of languages (Yip, 2002), few studies have investigated imitation of non-native tones. As with non-native consonants and vowels, perception of non-native tones is constrained by native phonological and phonetic factors (Chen, Best, & Antoniou, 2020; Reid, Burnham, Kassisopa, Reilly, Attina, Rattanasone, & Best, 2015; So & Best, 2010a, 2010b). Given that imitation depends on perception of the target items, non-native imitation could be affected by the perceptual influence of native categories. On the other hand, it can bypass some aspects of phonological encoding relative to non-native identification (Hao & de Jong, 2016). It is, therefore, unresolved whether and/or how non-native tone imitation is affected by native phonological and phonetic influences. To investigate this issue, the relationship between imitation of non-native tones and their perceived phonological and phonetic similarity to native tones must be examined with speakers of other tone languages, because direct assimilation to native phonological categories is not possible in non-tone languages as they lack tones at the segmental phonological tier (Best, 2019).

However, most previous studies on non-native tone production/imitation have been conducted with non-tone language speakers. Hao and de Jong (2016) found that English-native intermediate Mandarin learners imitated Mandarin tones more accurately than they identified them in Pinyin. This may imply that imitation can bypass some aspects of native phonological constraints. However, since English is a non-tone language, by definition there are no native segmental-level phonological biases that could affect performance on lexical tones (Best, 2019). Moreover, as the study did not collect data on perceptual assimilation to the native language, it also failed to address the issue of native language phonological and phonetic impact on non-native tone imitation.

Another study found that native speakers of a tone language, Cantonese, who had learned Mandarin as an L2, were also better at imitating than identifying Mandarin tones (Hao, 2012). Moreover, the correlation between their tone identification and imitation was not significant, suggesting that tone imitation does indeed bypass some aspects of native phonological constraints. According to PAM principles though, perception of the falling-rising tone should be affected by the native Cantonese tone system, which has a falling tone but lacks a falling-rising one. Indeed, they perceptually assimilated Mandarin falling-rising tone into Cantonese low falling tone, and inaccurately imitated its final rise, as verified by native Mandarin listeners. However, the participants had varying degrees of L2 Mandarin proficiency, and none were Mandarin-naïve, which confounds the L1 versus L2 influences on the relationship between perceptual assimilation and imitation. Imitation is likely to be affected more by the L1 in naïve imitators than in high proficiency L2 learners because naïve participants will necessarily refer to their native language when imitating the new language. Therefore, these results leave a number of non-native tone imitation issues unresolved.

In order to further examine how native language phonological and phonetic factors affect non-native tone imitation, it is essential to test both assimilation and imitation with native

speakers of a tone language who are naïve to the tones in the stimulus language. Imitators who are native tone language speakers can assimilate non-native tones into native tone categories and thus are susceptible to influences of their native tones at both phonological and phonetic levels. Being naïve to the stimulus language ensures that native language influence is maximised and experience with the target language is minimised. Comparing imitators from native languages that differ in tone systems, conversely, would allow observation of different language-specific native phonological and phonetic effects on imitation of the same set of non-native tones. The present study addresses these points by examining the phonological and phonetic influences of the native tone systems of Mandarin and Vietnamese on their speakers' non-native imitation of the tones of Thai, a third language unfamiliar to both groups.

Although SLM (Flege, 1995; SLM-r: Flege & Bohn, 2021) can be used to predict imitation based on similarities between native and non-native tones, that model focuses solely on phonetic categories and does not explicate the influence of the native language at both phonetic and phonological levels. Thus, we extended the core principles of PAM (Best, 1995; Faris, Best, & Tyler, 2018), which addresses both phonetic and phonological levels, to make predictions about imitation performance based on evidence of perceived similarity between native and non-native tones (Chen, Best, Antoniou, & Kasisopa, 2019; Chen, Antoniou, & Best, 2023) with the same target language and the same participants as in the present study. We disentangled native phonological influences as reflected in type of assimilation, that is, Categorised versus UnCategorised, and native phonetic influences as reflected in relative percentage choice and goodness ratings for a given native tone category (Chen et al., 2020) in making predictions about perceptual influences on non-native tone imitation.

Phonological influence is predicted to be strong for Categorised assimilations and weaker for UnCategorised assimilations. Within UnCategorised assimilations, the phonological influence is moderate for UnCategorised$_{\text{focalised}}$ assimilations, in which the non-native phone is assimilated as primarily similar to a single native category but choices of that native phoneme fall below the defined categorisation threshold (Faris et al., 2018). In UnCategorised$_{\text{clustered}}$ assimilation, the general native phonological influence is predicted to be weaker because the non-native phone is assimilated to a small set of native categories, which are all below the Categorised threshold but above chance level and thus none of them have unique or strong influence on assimilation. The native phonological influence is very weak for the UnCategorised$_{\text{dispersed}}$ assimilation, because assimilations of the non-native phone category are spread across many L1 response categories, all below chance level (Faris et al., 2018).

Within those phonological constraints, participants are nevertheless predicted to retain some residual sensitivity to within-category *phonetic* deviations of the non-native phones from their native categories. Residual native phonetic sensitivity is determined separately based on percent choice and goodness ratings of the chosen categories in the assimilation task.

1.2. A dynamic view of non-native speech processing for imitation

Non-native imitation performance can be strongly constrained by native phonology or accurately reflect the phonetic details of stimuli or reflect an interaction of the two, thus demanding a dynamic theoretical account of the imitation process. The Automatic Selective Perception model (ASP: Strange, 2011), which primarily accounts for performance variations in speech perception, can be extended to predict variations in imitation because perception provides the input for imitation. ASP claims that selective perception routines are used to process both native and non-native speech, as activated by the perceiver's detection of task-relevant information. When a task requires attention to phonetic differences that are essential to lexical distinctions, such as recognition of minimal contrasts, or phonological distinctiveness (Best, 2015; Best, Tyler, Gooding, Orlando, & Quann, 2009), and detection of phonological structures (phonemes, words, etc.), the activated routines constitute the phonological mode of speech perception, in which phonetic variations within a native phonological category (and non-native deviations from it) are likely to be unattended. On the other hand, when the task requires listeners to attend to fine-grained non-contrastive phonetic details (e.g., those that distinguish among accents or talkers), the phonetic mode is activated, allowing detection of phonetic variation within native phonological categories and of non-native deviations from native exemplars (see also Asano, 2018; Werker & Tees, 1984; Werker & Logan, 1985).

Perception precedes production in the process of imitation and therefore factors that affect selective perception routines should also impact imitation. Following this logic and extrapolating principles from ASP, we postulate that there are two modes of imitation that are linked to the two modes of selective perception. The dynamic balance between the phonological and phonetic modes in perception, and thus in imitation, is predicted to be influenced by cognitive factors such as memory load and stimulus variability (e.g., varying the number of talkers or vowel contexts) that can shift the balance of processing between abstract phonological categories and concrete fine-grained, non-contrastive phonetic details.

1.2.1. Memory load in imitation

The availability of phonetic details in short-term memory affects participants' ability to accurately imitate non-native stimuli. Imitators can only retain the rich array of fine-grained phonetic details in short-term memory for a limited time before they rapidly decay (Baddeley, 2010; Baddeley & Hitch, 1974). The longer the interval between the presentation of the stimuli and the imitation, the more likely it is that memory of the full range of phonetic details will fade. We will refer to the amount of time that participants must wait before imitating as memory load.

Non-native imitation should be phonetically more accurate under low memory load, that is, when the delay in imitating is brief and rich phonetic details of the target stimulus remain available, than under high memory load, when the delay is longer and phonetic details have decayed. Indeed, immediate imitation is reported to be more phonetically accurate and can bypass

native phonological constraints, relative to imitation delayed by a longer pause or an intervening task. When native speakers of Polish imitated English voiceless aspirated plosives [p^h, t^h, k^h], which are characterised by long-lag VOT values unlike Polish short-lag [p, t, k], their imitations were more English-like in the immediate condition (significantly longer VOTs), than when they instead had to read out a digit in the interval between the target item and imitating, delaying the imitation and interfering with retention of phonetic details, which significantly impaired their phonetic accuracy (Rojczyk, 2012). Auditory memory decays quickly, unlike the more sustained memory of abstract "encoded" phonological categories that are constrained by participants' native phonological systems. The findings suggest that participants must rely on their longer-lasting but phonetically-impoverished and native-biased phonological memory for delayed imitation.

However, even under low memory load, participants in some studies have failed to imitate non-native phones/features accurately. For example, when asked to imitate Japanese gemination contrasts, native speakers of German (which lacks gemination contrasts) deviated greatly from native Japanese speakers, and memory load had no effect (Asano & Braum, 2016). This implies that native phonological constraints on imitation can operate even under low memory load. Thus, it remains unresolved whether and how memory load will affect imitation of non-native lexical tones by native speakers of other tone languages.

1.2.2. Talker and vowel context variability

Talker variability, which in part reflects physiological and biomechanical differences in speakers' vocal tracts, has been reported to affect speech perception (Nusbaum & Morin, 1992). Extending the principles of ASP, we posited that high talker variability should bias listeners to use a more phonological mode of perception and detect more abstract information in the speech rather than low level phonetic information which varies across talkers. On the other hand, low talker variability should shift listeners to a more phonetic mode of perception because the phonetic level details in the speech are more nearly constant, allowing listeners to focus attention on less variable and more reliable within-talker phonetic details in duration, mean F0 or F0 maximum to minimum excursion. Evidence suggests that stimuli with high talker variability do indeed bias toward a native phonological mode of perception and result in lower accuracy than those with lower talker variability, even in native tone identification by Cantonese listeners (Wong & Diehl, 2003), as well as in discrimination of non-native Thai tones (Chen et al., 2019; Chen et al., 2023).

As imitation depends on both perception and production, we hypothesise that when talker variability in the Thai target stimuli is high, non-native imitators will shift toward a more phonological mode of perception and will be less sensitive to phonetic details. Consequently, their non-native imitations will be influenced by their native phonological perceptual routines and thus phonetically less accurate. When talker variability is low, however, imitators will

shift toward a more phonetic mode of perception and be more sensitive to phonetic details. As a result, their imitations will be less susceptible to native phonological constraints and phonetically more accurate.

Vowel context variability can also affect speech perception. Judging tones in variable vowel contexts reduces discrimination accuracy relative to when the vowel environment of the tones being judged is constant (Chen et al., 2019; Chen et al., 2023). We hypothesise that high vowel variability biases listeners to a more phonological mode of perception because the low-level phonetic information is variable, pushing listeners to instead rely on more abstract phonological information. On the other hand, we posit that low vowel variability biases listeners to a more phonetic mode of perception as the phonetic details are less variable and more reliable.

As with the proposed effect of talker variability on tone imitation, we expect that with variable vowel contexts, non-native imitators will be less sensitive to phonetic details and more affected by native phonological perceptual routines. Imitation in this case will be phonetically less accurate and more biased toward phonological categories of the L1 native language. With constant vowel contexts, on the other hand, we expect non-native imitators to be more sensitive to phonetic details and less affected by native phonological constraints. Consequently, their imitation will be more phonetically accurate. Few studies have examined the effects of talker and vowel context variability on non-native speech imitation, none of them addressing tone imitation. The present study was designed to test the above hypotheses.

1.3. The present study

The present study examined how Thai-naïve native Mandarin and native Southern Vietnamese (hereafter Vietnamese) participants' native tone systems influence their imitation of non-native Thai tones, and how this influence is modulated by memory load and stimulus variability. Many studies on tone perception and production have used Mandarin tone stimuli. We selected Thai tones as the stimuli to examine whether previous non-native imitation findings with Mandarin tones can be extended to another language from a different language family. It is also less likely to be familiar to speakers of other tone languages than is Mandarin.

Mandarin and Vietnamese listeners were recruited as participants because their native tone systems contain both level and contour tones, like Thai, and yet both systems differ from that of Thai as well as from each other. Thus, PAM predicts they will perceptually assimilate non-native Thai tones differently into their native tone categories phonologically as reflected in Categorised or UnCategorised types, and phonetically as reflected in percent choices and goodness ratings. To evaluate imitation performance, we compared key acoustic measures of the imitations with the target stimuli. The less the imitations deviate from the target stimuli, the more closely the phonetic details of the targets have been imitated.

In order to make predictions for imitation performance in consideration of native language constraints, we extrapolated from the principles of PAM/PAM-L2 (Best, 1995; Best & Tyler, 2007; Chen et al., 2020) and disentangled native phonological versus phonetic contributions to non-native tone categorisation and imitation by considering the type of assimilation for phonological influence and the relative percentage choice and goodness ratings for phonetic influence. Phonological influences on non-native tone perception are predicted to be strong for Categorised assimilation, which should result in imitation productions that are more like the participants' native tones, and weaker for UnCategorised assimilation. Within those phonological constraints, strong residual perceptual sensitivity to within-category phonetic variations of the non-native phones from the participants' native tone categories should be indicated by low percent choice and goodness ratings, whereas weak residual perceptual sensitivity to such phonetic variations should be reflected in high percent choice and goodness ratings. Strong residual phonetic sensitivity in perception should facilitate more accurate imitation of non-native tones.

Moreover, we predicted that imitations would be more phonetically accurate (i.e., less deviant from the target Thai stimulus details) under low memory load when fine phonetic details are available in short term memory and imitators can engage in a more phonetic mode of perception according to the principles we have extrapolated from ASP. In contrast, under high memory load, phonetic details of the target item were expected to have faded from short memory, such that imitators should engage in a more phonological mode of perception. Thus, we proposed that imitation would be more constrained by native language phonological influences when memory load is high. For tones assimilated into a native category that deviates phonetically from the target tone, imitations should be deviant as well if the phonological mode is strongly engaged.

In addition, when talkers or vowels vary within a test block, requiring participants to process linguistically irrelevant phonetic differences in parallel with the crucial tone-related details, we predicted that imitators should engage in a more phonological mode of perception according to ASP principles. As a result, their imitations should be phonetically less accurate and more deviant from the target stimuli. In contrast, when the talker and vowel within a block are constant, we posit that imitators can engage in a more phonetic mode of perception and their imitation should be more accurate, deviating less from the target stimuli.

2. General method

2.1. Lexical tones in Thai, Mandarin, and Vietnamese

Thai, Mandarin, and Vietnamese differ in the number and types of tones in their native inventories. We used Chao values (Chao, 1930), in which F0 height at tone onset and offset and sometimes at an intervening point is referenced by numbers 1–5 ranging from low to high, as a priori phonetic descriptions of the tones in each language. Here, Thai tones are designated with T, Mandarin

tones with M, and Vietnamese tones with V, and we describe *phonological* feature distinctions in terms of perceived abstract pitch heights (i.e., high, mid, low) and contours (i.e., level, rising, falling, falling-rising), as opposed to the *phonetically* specific, concrete F0 properties that Chao numbers are intended to capture.

Thai, the target language, has three phonologically level tones, high-level T45, mid-level T33, low-level T21; and two contour tones, rising T315, and falling T241 (Reid et al., 2015). Mandarin, on the other hand, has four tones: Level M55, rising M35, falling-rising M214, and falling M51 (Yip, 2002). The Vietnamese imitators in the present study all spoke the Southern dialect, which has five tones: Two phonologically level tones, high-level V44 (ngang), low-level V22 (huyền); and three contour tones, rising V35 (sắc), falling V21 (nặng), and falling-rising V214 (Nhàn, 1984). V214 instantiates the South Vietnamese merger of two Northern/standard dialect tones, V214 (hỏi), and V415 (ngã) (Brunelle, 2009; Chen et al., 2020).

In addition to phonological and phonetic descriptions of lexical tones in the three languages, we also present an acoustic analysis of these tones.

2.2. Acoustic analyses of tones in each language and in imitations

The same acoustic processing procedures were applied in extracting F0 from the Thai tones (as presented in 2.2 and 2.3) and from participants' native tones (as in 2.2), as well as for their imitations of Thai tones (as in Experiment 1 and 2). First, *ProsodyPro* (Xu, 2013), a *Praat* (Boersma, 2001) script, was used to extract the F0 contours of the stimuli and their imitations at 10 time-normalised points of F0. In order to make F0 values comparable across different speakers, all F0 values were then Lobanov-normalised (Lobanov, 1971), which reflects how much the mean F0 for a given data point varies from the F0 mean of the speaker. We used the most stable portion of the contour (i.e., between 10% to 90% of the syllable length) to plot all tone contours and calculated F0-related measures.

Time- and Lobanov-normalised (Lobanov, 1971) mean F0 trajectories of the tones in each language are presented in **Figure 1**. The Thai syllables were recorded for a separate study (Burnham, Kuratate, McBride-Chang, & Mattock, 2009), from four female Thai speakers (M_{age} = 30.3 years, *SD* = 3.8 years) who were all born and raised in Bangkok, and are used here with permission from the authors. They were recorded in citation form in two syllables (/ma:/ and /mi:/),[1] in a sound-treated booth at the MARCS Institute for Brain Behaviour and Development, Western Sydney University, using a Lavalier AKG C417 PP microphone with the sampling rate of 48 kHz and 16-bit

[1] The main reason we used long Thai vowels is that the Thai language has vowel length contrasts, and only long vowels allow all five tones to distinguish real words that have the identical consonant-vowel structure and differ only by lexical tones. Short vowels in Thai have different tone distribution and phonotactic constraints than long vowels. Most short vowels allow only three (low, falling, and high) lexical tones instead of five tones, and syllables with short vowels are normally considered to have glottal stops as final consonants (i.e., they are closed or checked, not open syllables as we needed for this study).

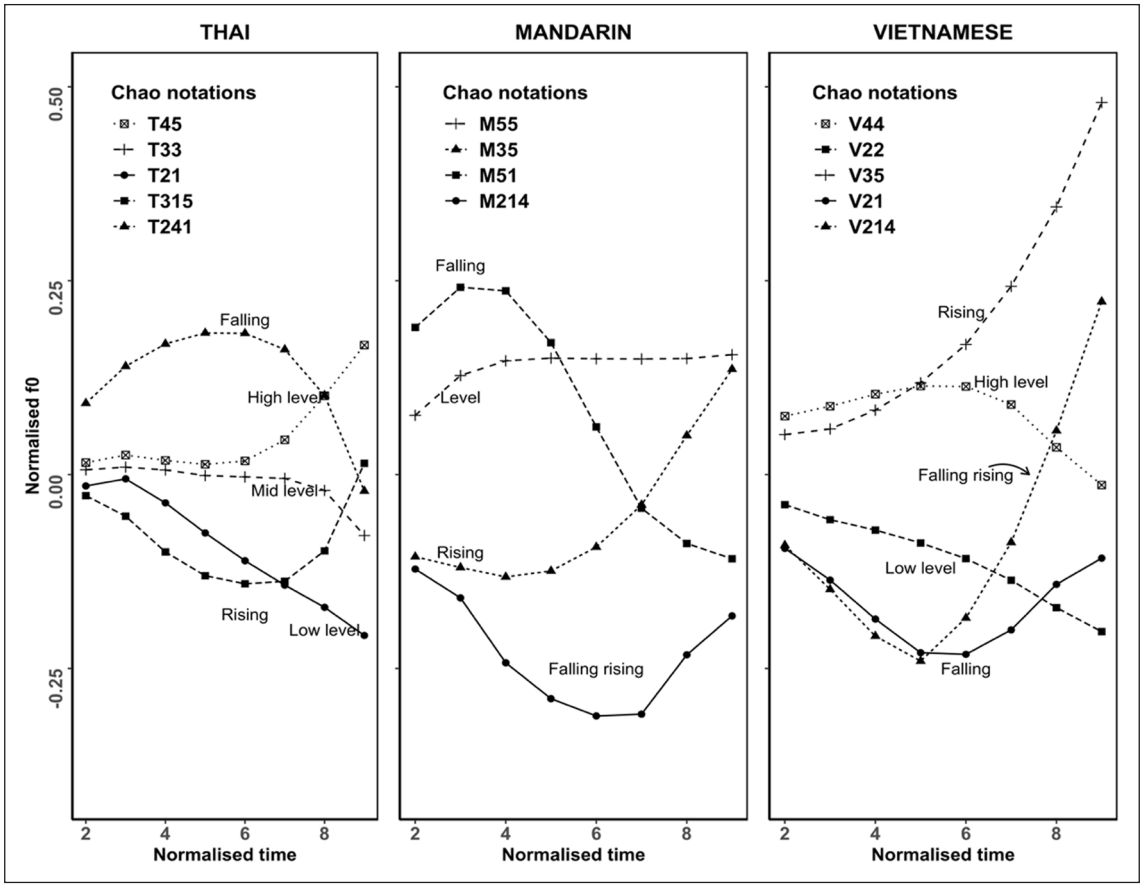

Figure 1: Time- and Lobanov-normalised (Lobanov, 1971) F0 contours of Thai, Mandarin, and Southern Vietnamese tones.[3] Note that concrete phonetic contours of actual tone tokens may differ from their designated abstract phonological characterisation, particularly in the case of phonologically level tones.

resolution. Note that the Thai database we used provides different number of tokens from each participant: Five tokens per target tone per syllable from two participants, four tokens each from the third participant, and six each from the fourth participant, yielding 20 tokens per target tone per syllable (20 tokens × 5 tones × 2 syllables = 200 in total).

Each native tone of the imitator groups' languages was produced by four female native speakers of each language (Mandarin, M_{age} = 27.0 years, SD = 2.2 years; Vietnamese, M_{age} = 21.0 years, SD = 3.0 years) in the syllables /ma/ and /mi/,[2] eight times each by each speaker,

[2] Note that neither Mandarin nor Vietnamese use contrasting vowel length, so we cannot equate vowel length between the Thai stimuli, where a length value must be chosen, and the imitator languages, for which there are no vowel length distinctions.

[3] In Chen et al. (2020), we had asked the Southern Vietnamese speakers to produce both V214 (hỏi) and V415 (ngã), but consistent with the reports of merger they showed no significant acoustic differences. So here they were averaged and labelled as the single South Vietnamese phonologically falling-rising tone V214.

in random order in a prior study (Chen et al., 2020). Mandarin items were elicited via Pinyin, Vietnamese items via the orthography of their language. Thus, there were 64 tokens (2 syllables /ma/ and /mi/ × 4 tones each × 8 repetitions) for each Mandarin informant and 96 tokens (2 syllables /ma/ and /mi/ × 6 tones each[1] × 8 repetitions) for each Vietnamese informant. Mandarin and Vietnamese productions were recorded using a Zoom H4n digital speech recorder with a sampling rate of 44.1 kHz and 16-bit resolution in a quiet testing booth at The MARCS Institute, Western Sydney University. All target syllables were meaningful morphemes in the respective languages (for English glosses, see Table D1-3 in the Appendix D).

Given that average F0, direction of F0 change, duration of tone, location of highest and lowest F0 values, and slope of F0 rise/fall are reported to be the primary factors affecting the perception of lexical tones (Gandour, 1978), we selected four acoustic measures to characterise tones and their imitations for the present study: Duration, $F0_{mean}$, F0 maximum to minimum excursion ($F0_{excursion}$), and F0 maximum location ($F0_{maxloc}$, the timepoint of the maximum F0 along the eight time-normalised points divided by eight) (see Appendix A, Table A.1 in the supplementary materials). $F0_{excursion}$ distinguishes level tones from contour tones, and contour tones such as T241 and T315 can be differentiated by $F0_{maxloc}$. In a Principal Component Analysis of lexical tones (Chen, Best, Antoniou, & Kasisopa, 2018), these acoustic measures outweighed other measures in differentiating Thai, Mandarin, Southern and Northern Vietnamese tones. To make more concrete predictions about the direction of deviations in imitation, we calculated the 95% confidence intervals for the four acoustic measures to compare Thai tones with Mandarin and Vietnamese tones (see **Figure 2**). Note that Thai long vowels were compared with Mandarin and Vietnamese vowels, which do not have contrastive length differences; they are neither short nor long. The duration values were not normalized, thus allowing us to determine native language influences on vowel duration in imitations.

2.3. Thai stimulus materials

Two tokens of each Thai target item produced by two native speakers (27 and 33 years old), out of the four used in 2.2, were selected (see **Figure 3**) as imitation stimuli. These tokens were judged to be the most natural sounding by a third native Thai listener. For the stimulus variability factors, we systematically manipulated the talker and vowel variability of the Thai target stimuli (i.e., in constant versus variable blocks). **Figure 3** shows the talker and vowel variability in the Thai stimuli in terms of their F0 contours. Raw F0 was used here, rather than normalised F0 values, to provide the variations contained in the stimuli and give readers an idea of the true F0 variations that imitators actually faced in variable talker conditions. It should be noted that the variability among talkers and vowels is much greater for T241 than the other Thai target stimuli. This could reduce imitation accuracy of T241 relative to the other Thai tones in the variable talker blocks.

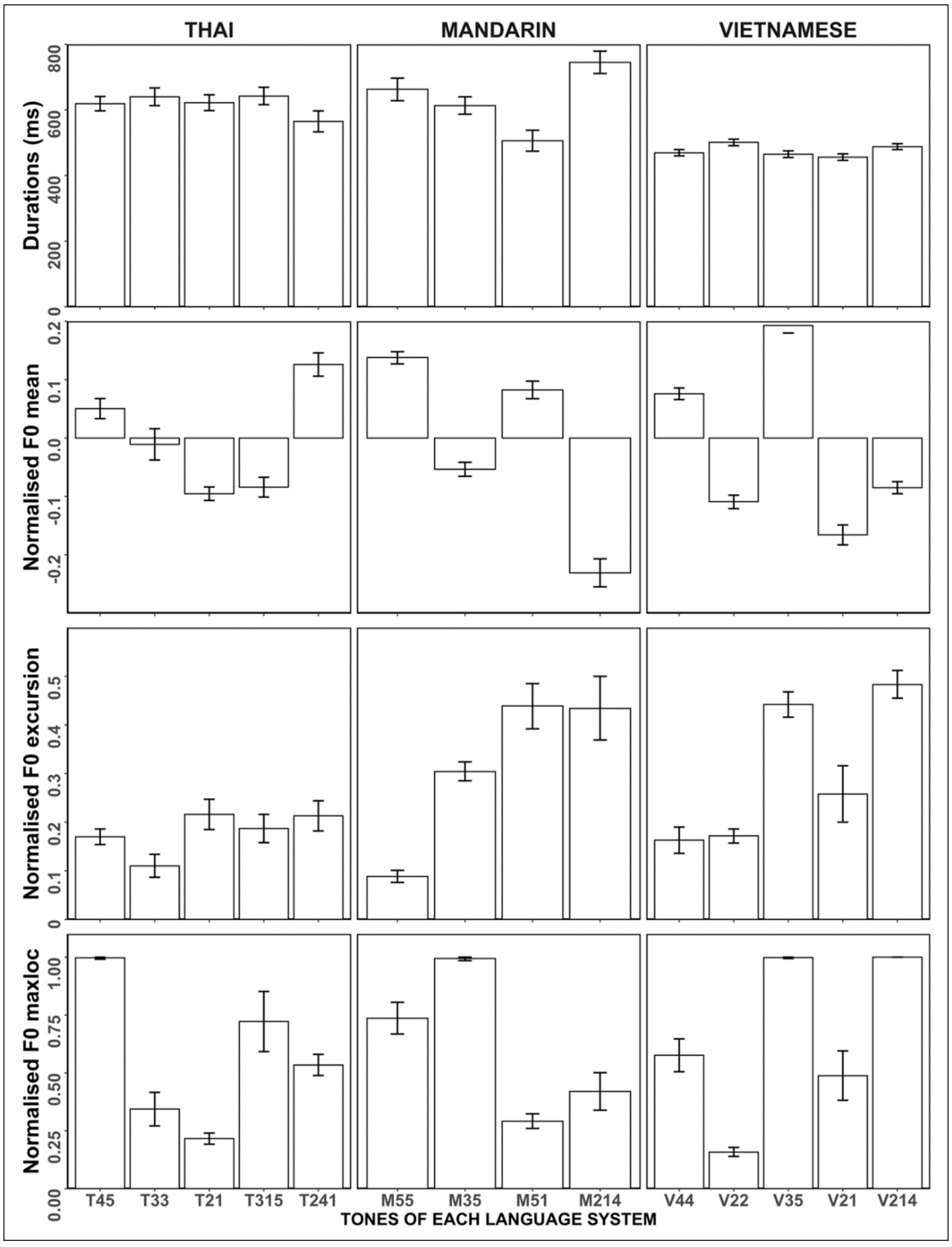

Figure 2: Acoustic measures of tones in Thai (40 tokens per tone), Mandarin, and Vietnamese (64 tokens per tone). F0mean, F0excursion are Lobanov-normalised Hz values. The error bars indicate the 95% confidence intervals. F0maxloc values represent the location of the maximum F0 in the time-normalised contour.

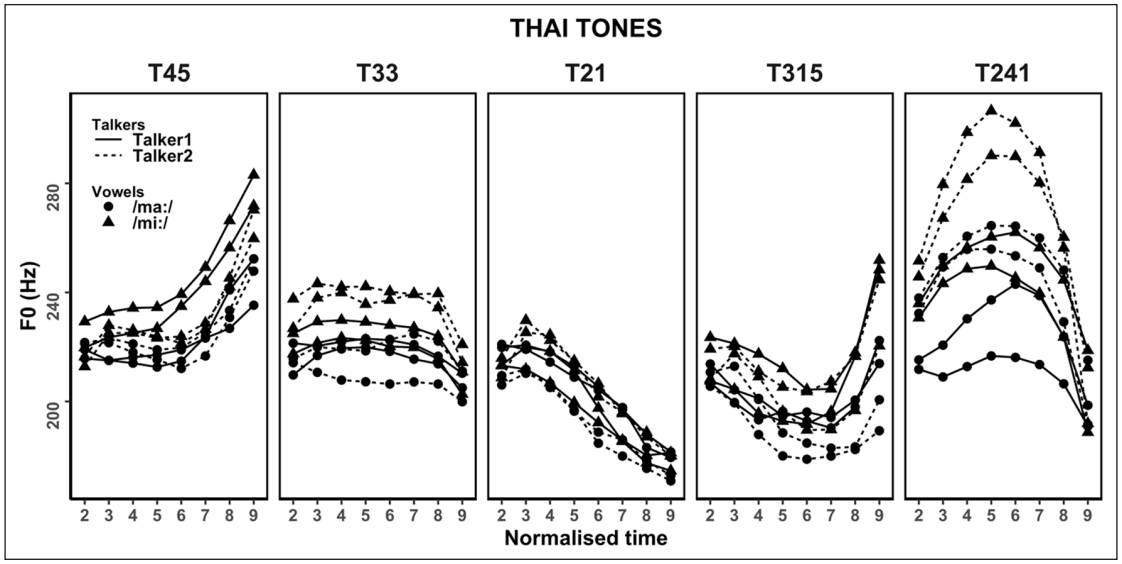

Figure 3: Talker and vowel variability in mean F0 contours of Thai stimuli (time-normalised: 2 tokens × 2 talkers × 2 syllables × 5 tones).

2.4. Procedure

In the current imitation study, memory load was operationalised as the time between the end of the stimulus and the signal for participants to produce their imitation, referred to as the imitation interval. When selecting the intervals, some previous studies used 0 versus 2500 ms in consonant imitation (Asano & Braum, 2016), whereas other used 500 versus 1500 ms in consonant perception (Werker & Tees, 1984; Werker & Logan, 1985). We used 500 versus 2000 ms to maximumly shift imitators between two modes of imitation and at the same time keep the experiment duration feasible. An additional reason for this decision is to align this study to the other perception experiments in the same larger project (Chen et al., 2023), in which we used the same intervals, so that perceptual assimilation results could be used to form predictions in the present study. Under low memory load, a message "Imitate now!" appeared 500 ms after the offset of the stimulus to alert participants to imitate. Under high memory load, the same message was shown 2000 ms after the offset of the stimulus. Under both memory load conditions, the inter-trial interval was one second. We also blocked talker variability (constant = one talker vs. variable = two talkers) and vowel variability (constant = /ma:/ or /mi:/ vs. variable vowels = /ma:/ and /mi:/) across the experiment.

Participants were instructed to imitate the tones in the stimuli as faithfully as possible after they heard the auditory stimulus and saw the starting signal. They were not asked to imitate talker-related characteristics. We note that our participants may have attended to, and possibly reproduced, voice quality in imitating some Thai stimuli given that it is arguably of linguistic importance for some tones in Mandarin and Vietnamese. However, the focus of the present study is on the F0 patterns (height, contour, etc.) in the imitations relative to the Thai targets.

Before the test session, participants completed 10 practise trials. Then each participant completed 160 imitation trials (2 syllables × 5 tones × 2 tokens × 2 talker variability conditions × 2 vowel variability conditions × 2 repeats) in total.

Participants were tested individually in testing booths at Western Sydney University, the University of New South Wales, and Macquarie University. Stimuli were presented through Sennheiser HD 280 Pro headphones at 72 dB SPL from a Dell Latitude 7280 laptop running E-Prime Professional 2. Participants' responses were recorded on a portable digital speech recorder (ZOOM H4n) at 44.1 kHz sampling rate and 16-bit resolution.

2.5. Acoustic evaluation of imitations

For evaluating non-native tone imitation, we selected duration, $F0_{mean}$, $F0_{excursion}$, $F0_{maxloc}$ as in 2.2. In addition, to quantify the acoustic deviations of the imitations from their target Thai stimuli for use in statistical analyses, we calculated deviation scores following the procedure in Wang, Jongman, and Sereno (2003) by subtracting the values of duration and the three Lobanov-normalised F0-related acoustic measures for the target stimulus token from the values for the imitation of that target token. We used signed deviation scores for the statistical modelling instead of absolute values so as to take the direction of deviation into consideration. This is essential to our study as we are interested in, and do not assume which direction of effects, the manipulation of memory load and stimulus variability will have on different acoustic measures for each tone.

3. Experiment 1: Imitation of Thai tones by Mandarin speakers

3.1. Participants

Native speakers of Mandarin ($n = 32$), who were living in Sydney to attend university or to work, participated in the experiment. All had participated in a related study on perception of Thai tones (i.e., Chen et al., 2023), which may constitute prior Thai experience of a very limited nature. They did the second session of the perceptual experiment before this imitation task on the same day. They were divided into two groups for each imitation condition (low memory load: $M_{age} = 26.6$ years, $SD = 7$ years, 10 females; high memory load: $M_{age} = 26.5$ years, $SD = 6.4$ years, 10 females). Although the stimuli were female utterances of Thai tones, both male and female participants were required to imitate the perceived form of the lexical tone, rather than the exact acoustic F0 values. In addition, from a developmental perspective, male infants can naturally and automatically imitate the global form of their mothers' and fathers' words and phrases, including tones, when acquiring their native language, despite the even much larger difference in their F0s and other formant values, relative to their parents.

Participants completed a background questionnaire before the test. The Mandarin-speaking participants were all born and raised in China (various regions and native dialects:[4] Tianjin, Anhui, Sichuan, Shandong, Henan, Hunan, Jilin, Jiangsu, Jiangxi, Shanxi, Xinjiang, Shanghai, Inner Mongolia); all were educated in Mandarin from early childhood through high school, and they used Mandarin on a daily basis. They had also learned English before coming to Australia and the average length of residence in Sydney was 1.5 years (SD = 1.6 years, with one missing data point). Note that their length of residence in Sydney, which is an English-speaking community, and their experience with English would not modulate their native tone phonological knowledge, as English is not a tone language. None had ever lived in Thailand or learned Thai or had more than two years of formal musical training, which is known to facilitate tone perception and imitation (Gottfried, Staby, & Ziemer, 2004). All reported normal hearing. The experiments were approved by the Western Sydney University Human Research Ethics Committee (H12560) and all participants signed a consent form prior to testing and were compensated for their time (AU$20).

3.2. Predictions

Extending PAM principles to imitation, we argue that if a non-native tone is Categorised as a native tone, then the imitation of that tone will be similar to the imitators' native tone. In Chen et al. (2023), the same Mandarin participants perceptually assimilated the present Thai tone stimuli into their four native tone categories under low and high memory loads (see **Table 1**). We used those perceptual data as the basis for making predictions about native phonological and phonetic influences on imitations of Thai tones. To quantify residual phonetic sensitivity from the prior perceptual data, we first divided percent choice of the native tones above chance (25% given their four-tone system) into three ranges: Low, Medium, and High. These ranges reflect strong, moderate, and weak residual phonetic effects, respectively. Low spanned 25%–49% of choices, Medium 50%–75%, and High 76%–100%. For the category-goodness ratings, we also divided the scale into three ranges: Low = 1.0–2.9, Medium = 3.0–4.9, and High = 5.0–7.0. These ranges reflect strong, moderate, and weak residual phonetic effects, respectively.

Under both memory loads (Chen et al., 2023), T33 was Categorised as M55, T45 was Categorised as M35, and T241 was Categorised as M51 (see **Table 1**). Both percent choice and goodness ratings for the three assimilations were in the medium-to-high range, suggesting relatively low residual phonetic sensitivity to differences between native and non-native tones.

[4] In future research, it would be desirable to form more homogeneous groups of participants and have better control of their dialects (e.g., recruiting only Beijing Mandarin speakers). For practical reasons, strict control of dialect background of Mandarin participants living in Sydney is difficult, if not impossible. With this said, the possible effect of dialect background differences adds potential variation to the assimilation patterns. In this experiment, we used participants' assimilation patterns to directly predict and account for variations in imitations, rather than using phonetic descriptions of any Mandarin dialect variations. In this way, any dialect effects on perception of the non-native Thai tones should be consistent across assimilation and imitation tasks.

Thai stimulus		T45		T33		T21		T315		T241	
	Response	%	rating	%	rating	%	rating	%	rating	%	rating
Low memory load	M55	–	–	**77.3***	5.3	19.9	3.3	–	–	21.6	4.2
	M35	**88.8***	5.8	2.8	2.8	1.6	2.1	**48.6**	5.5	2.9	3.7
	M51	–	–	19.2	4.9	**52.7***	4.6	–	–	**75.1***	5.6
	M214	10.3	4.3	–	–	25.8	3.7	**51.2**	5.4	–	–
Assimilation		C		C		C		$U_{clustered}$		C	
	M55	–	–	**84.7***	5.2	26.4	3.2	–	–	**28.1**	5.0
High memory load	M35	**85***	5.3	–	–	1.1	3.3	**44.2**	5.1	–	–
	M51	–	–	13.4	4.4	**66.2***	4.0	–	–	**71.2***	5.1
	M214	14.7	4.8	1.7	5.9	6.3	3.8	**55.6**	5.5	–	–
Assimilation		C		C		C		$U_{clustered}$		C	

Table 1: Assimilation of Thai tones into Mandarin tone categories under low versus high memory loads (data from Chen et al., 2023). Categories in bold are choices that were significantly above chance: 25% for Mandarin; "*" = Categorised tone. Assimilations: C = Categorised, U = UnCategorised. Rating: 1 = poor, 7 = Perfect; mean ratings are displayed. "–" = responses <1%.

T21 was Categorised as M51 under both memory loads but with percent choice and ratings in the medium range, suggesting moderate residual phonetic sensitivity to differences between native and non-native tones. T315 was an $UnCategorised_{clustered}$ assimilation under both memory loads and was split between M35 and M214. For $UnCategorised_{clustered}$ assimilations, we predicted the native phonological influence would be relatively weak. Percent choice of native response categories for T315 were in low and/or medium ranges, whereas goodness ratings were in the high range, suggesting moderate residual phonetic sensitivity to differences between T315 and the chosen native tones, which is expected to moderately facilitate accurate imitation of non-native tones.

3.3. Results

There were 24 missing data points (0.4%) for the Mandarin group, where participants started the imitation and self-repaired in the middle and we excluded these data from the analysis. The signed deviation scores for duration, $F0_{excursion}$, $F0_{mean}$, and $F0_{maxloc}$ were each selected as dependent variables and fitted with a separate linear mixed-effects model using *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in *R* (R Core Team, 2018) (for deviation scores of imitations by Mandarin participants, see Appendix A, Table A.2). Thai tone (T45, T33, T21, T315, T241), memory load (low vs. high), and talker (constant vs. variable) and vowel variability

(constant vs. variable) were used as fixed factors. Four models were built to test all main effects and interactions. For each model, we first ran the analysis with participants as a random factor including random slopes for all within-subject factors, namely talker and vowel variability (as suggested by Barr, Levy, Scheepers, & Tily, 2013). The models converged but were too complex to estimate *p* values for the fixed effects of interest using the Kenward-Roger degrees of freedom approximation (Halekoh & Hojsgaard, 2014). Thus, we dropped the random slopes and participants were specified as a random intercept in each model. Here we report main effects for multilevel factors or interactions using *F*-tests via the *Anova* function from the *car* package (Fox & Weisberg, 2019) in R (see **Table 2**), rather than comparisons with the baseline level using *t*-tests. Thus, the reported main/interaction effects were averaged across all levels of the other effects, and can be directly used to test our predictions.

There were significant main effects of tone for all four deviation scores, and of talker variability for $F0_{maxloc}$ deviation scores. However, the main effects of memory load and vowel variability were non-significant for all deviation scores. There were significant tone × talker variability interactions for duration, $F0_{mean}$, and $F0_{maxloc}$, and tone × memory load interactions for $F0_{mean}$, $F0_{excursion}$, and $F0_{maxloc}$.

To further examine the tone main effects, we ran pairwise multiple *t*-tests with Tukey adjustments for all tone differences for each deviation score (for statistical details see Appendix B, Table B1). All pairwise tone comparisons were significant for the duration deviation scores. Given that we used signed deviation scores in the model, the analysis did not explicitly compare the scores with 0, which corresponds to the native Thai value. When interpreting the results, we used the 95% confidence interval to evaluate the significance. If the confidence interval did not include 0, we interpret the relevant scores to be significantly either higher or lower than the native Thai value. T33 imitations had the largest deviation scores for syllable duration (M = 0.064, 95% CIs [0.057, 0.071]), thus being the least accurate. Those for T241 (M = 0.058, 95% CIs [0.051, 0.065]) were the second largest. Both T21 (M = –0.014, 95% CIs [–0.021, –0.006]) and T315 (M = –0.023, 95% CIs [–0.030, –0.015]) had negative deviation scores, indicating that the imitations were shorter than the target stimuli. The duration of T45 imitations (M = 0.007, 95% CIs [0.0004, 0.015]) was the most accurate, as indicated by the smallest absolute value of deviation scores.

All pairwise tone comparisons were also significant for $F0_{mean}$ deviation scores, which were positive only for T315 (M = 0.046, 95% CIs [0.042, 0.050]), indicating higher $F0_{mean}$ in the imitations than the target stimuli and negative for T45 (M = –0.021, 95% CIs [–0.025, –0.018]), T33 (M = –0.011, 95% CIs [–0.015, –0.008]), indicating lower $F0_{mean}$ for imitations than the target stimuli. As the confidence intervals of the scores for T241 (M = –0.037, 95% CIs [–0.042, 0.032]) and T21 (M = –0.0004, 95% CIs [–0.004, 0.03]) included 0 (i.e., a non-significant difference from the Thai target value), we interpret the relevant scores to be close to the native norm, indicating accuracy in $F0_{mean}$ imitation.

	Duration			$F0_{mean}$			$F0_{excursion}$			$F0_{maxloc}$		
	F	***df***	***p***	***F***	***df***	***p***	***F***	***df***	***p***	***F***	***df***	***p***
Tone	121.9	4, 5026	**<.001**	253.1	4, 5026	**<.001**	97.7	4, 5026	**<.001**	194.6	4, 5026	**<.001**
Talker (Tlk)	1.0	1, 5026	0.328	0.3	1, 5026	0.571	0.9	1, 5026	0.346	118.6	1, 5026	**<.001**
Vowel (V)	1.0	1, 5026	0.306	1.5	1, 5026	0.218	0.0	1, 5026	0.829	0.3	1, 5026	0.578
Memory (Mem)	0.0	1,30	0.989	4.1	1, 30	(0.052)	3.4	1, 30	(0.076)	0.3	1, 30	0.600
Tone × Tlk	11.7	4, 5026	**<.001**	9.9	4, 5026	**<.001**	1.3	4, 5026	0.250	50.8	4, 5026	**<.001**
Tone × V	0.6	4, 5026	0.649	0.7	4, 5026	0.584	1.0	4, 5026	0.390	0.5	4, 5026	0.754
Tone × Mem	1.6	4, 5026	0.168	13.8	4, 5026	**<.001**	8.1	4, 5026	**<.001**	2.8	4, 5026	**0.026**
Tlk × V	0.6	1, 5026	0.436	0.3	1, 5026	0.565	2.6	1, 5026	0.106	0.1	1, 5026	0.754
Tlk × Mem	0.1	1, 5026	0.787	0.1	1, 5026	0.705	0.5	1, 5026	0.471	0.3	1, 5026	0.609
V × Mem	0.6	1, 5026	0.444	0.0	1, 5026	0.941	1.0	1, 5026	0.318	0.1	1, 5026	0.716
Tlk × V × Mem	0.7	1, 5026	0.388	0.1	1, 5026	0.745	0.5	1, 5026	0.479	2.6	1, 5026	0.108
Tone × Tlk × V	0.2	4, 5026	0.959	0.3	4, 5026	0.878	0.9	4, 5026	0.468	0.5	4, 5026	0.717
Tone × Tlk × Mem	0.1	4, 5026	0.978	0.2	4, 5026	0.923	1.3	4, 5026	0.270	0.4	4, 5026	0.776
Tone × V × Mem	0.2	4, 5026	0.951	0.4	4, 5026	0.815	0.1	4, 5026	0.970	0.2	4, 5026	0.916
Tone × Tlk × V × Mem	1.0	4, 5026	0.426	0.1	4, 5026	0.970	0.6	4, 5026	0.631	0.5	4, 5026	0.720

Table 2: Model details of acoustic measure deviation scores of Mandarin imitators. Significant effects are in bold; marginal effects are in parentheses.

Again, all pairwise tone comparisons were significant for the $F0_{excursion}$ deviation scores except for the T33-T241 comparison. The scores were positive for all tones, indicating that Mandarin imitators generally enlarged the range of the F0 contour in the imitations (i.e., they hyper-articulated the contours). These scores were largest for T315 (M = 0.171, 95% CIs [0.161, 0.181]), followed by T33 (M = 0.114, 95% CIs [0.106, 0.123]), T241 (M = 0.102, 95% CIs [0.089, 0.114]), T45 (M = 0.079, 95% CIs [0.069, 0.088]), suggesting T315 was the least accurately imitated. T21 (M = 0.053, 95% CIs [0.045, 0.061]) was the most accurately imitated, with the smallest $F0_{excursion}$ deviation scores.

Finally, all pairwise tone comparisons were again significant for $F0_{maxloc}$ deviation scores. These scores were positive for T21 (M = 0.021, 95% CIs [0.012, 0.031]) and T315 (M = 0.063, 95% CIs [0.044, 0.082]), indicating that the F0 peak in imitation was delayed relative to that in the Thai target stimuli. Imitation of T315 was less accurate than T21. In contrast, T241 (M = –0.167, 95% CIs [–0.179, –0.156]), T45 (M = –0.068, 95% CIs [–0.075, –0.061]), T33 (M = –0.019, 95% CIs [–0.034, –0.003]) had negative $F0_{maxloc}$ values, indicating F0 peaks were realised earlier in imitations than in the target stimuli. T241 deviation scores were largest in absolute value, indicating imitation of $F0_{maxloc}$ was least accurate for this tone. T45 was imitated more accurately than T33 on this measure.

To further examine tone × memory load interactions for $F0_{mean}$, $F0_{excursion}$, and $F0_{maxloc}$ deviation scores, for each of those scores we conducted pairwise memory load condition comparisons for each tone using Tukey adjustments (for statistical details see Appendix B, Table B.2). $F0_{mean}$ deviation scores revealed more accurate imitation of T315 under low (M = 0.036, 95% CIs [0.030, 0.041]) than high memory load (M = 0.056, 95% CIs [0.051, 0.062]), but less accurate imitation of T45 under low (M = –0.033, 95% CIs [–0.038, –0.028]) than high memory load (M = –0.010, 95% CIs [–0.014, –0.005]). Although there were some variations in $F0_{excursion}$ and $F0_{maxloc}$ deviation scores between the two memory load conditions, none of these differences were significant for the same tone.

Unexpectedly, the main effect of talker variability in $F0_{maxloc}$ deviation scores reflects a pattern of overall more accurate imitation in variable (M = –0.003, 95% CIs [–0.013, 0.007]) than constant talker (M = –0.065, 95% CIs [–0.072, –0.058]) blocks, but this was mediated by a higher order tone × talker variability interaction. Thus, we ran multiple comparisons as above to break down the tone × talker variability interactions for duration, $F0_{mean}$, and $F0_{maxloc}$ deviation scores (see **Figure 4** for the general F0 contours, and for full statistical results see Appendix B, Table B.3). The crucial comparisons are those for the same tone between the two talker variability conditions (significant differences described below). The confidence intervals of T21 and T45 durations in the constant talker blocks spanned across zero (M_{T21} = 0.008, 95% CIs [–0.002, 0.018]; M_{T45} = –0.005, 95% CIs [–0.015, 0.005]), indicating that the mean was not significantly above or below the Thai target values (zero) and thus the imitation was accurate.

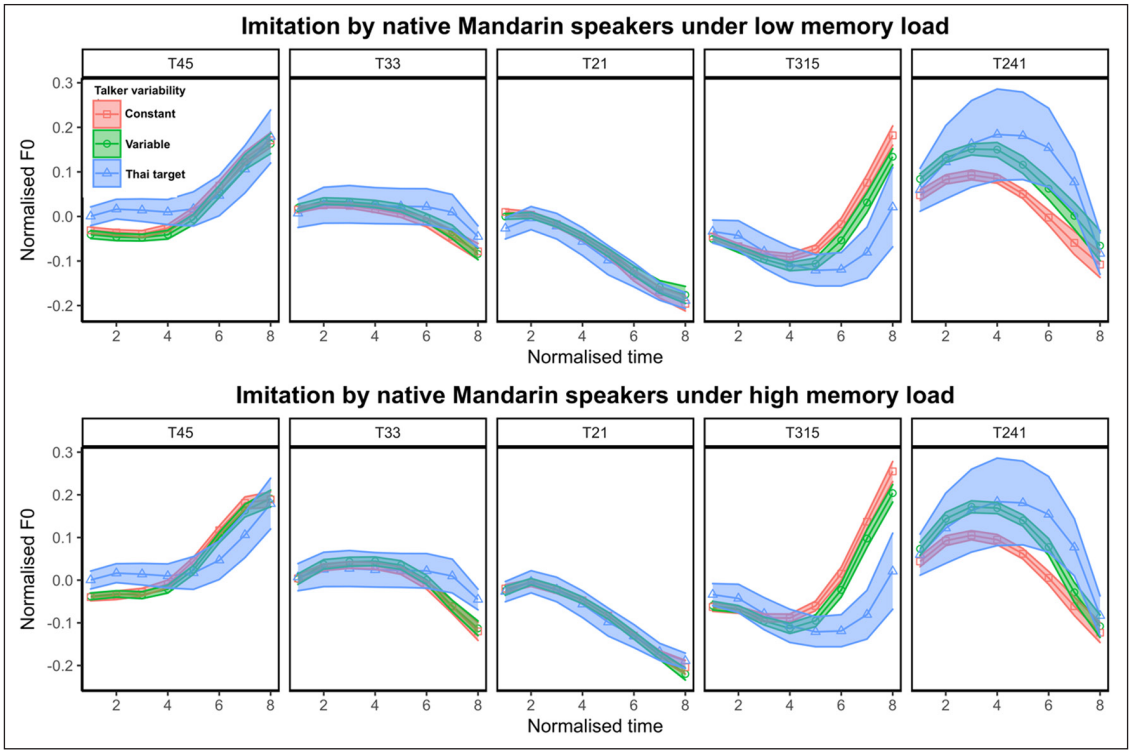

Figure 4: The time-and-Lobanov-normalised mean F0 contours of the Thai target stimulus tones and their imitations by Mandarin participants in the constant versus variable talker blocks. Ribbons indicate 95% confidence intervals for each mean contour type (blue = Thai stimuli; green = variable talker; red = constant talker).

In contrast, the confidence intervals of T21 and T45 in the variable talker blocks were on the left or right side of zero (M_{T21} = –0.035, 95% CIs [–0.046, –0.025]; M_{T45} = 0.020, 95% CIs [0.009, 0.030]), which indicates the mean was significantly below or above the Thai target value and thus the imitation was less accurate. Taken together, the T21 and T45 duration imitations were more accurate in constant than variable talker blocks.

Imitation of $F0_{mean}$ for T241 was also more accurate in constant (M = –0.026, 95% CIs [–0.031, –0.020]) than variable talker blocks (M = –0.048, 95% CIs [–0.056, –0.040]). However, for T241, imitation of $F0_{maxloc}$ was *less* accurate in constant (M_{T241} = –0.188, 95% CIs [–0.204, –0.172]) than in variable talker blocks (M_{T241} = –0.147, 95% CIs [–0.162, –0.131]). For T315 and T33 imitation of $F0_{maxloc}$, the directions of deviation were different across the two talker variability conditions (the peak was earlier in constant talker blocks: M_{T315} = –0.029, 95% CIs [–0.038, –0.019]) and M_{T33} = –0.081, 95% CIs [–0.102, –0.061]; but was delayed in variable talker blocks: M_{T315} = 0.155, SD = 95% CIs [0.119, 0.190] and M_{T33} = 0.044, 95% CIs [0.022, 0.066]). It should be noted that variations in $F0_{maxloc}$ may be more meaningful or salient for contour tones like T241 with the F0 maximum location in the middle of the syllable than they are for tones like

T33, which is a level tone, or for dipping contour tones like T315, whose F0 maximum is realised at the end. Deviations of $F0_{maxloc}$ in T33 and T315 across two talker variability conditions only indicate low level phonetic variations that do not reflect any phonological abstraction about the tone type.

3.4. Discussion

First, Mandarin participants imitated Thai tones with overall good accuracies particularly for T21, but the phonetic accuracy varied in strength for different tones in a way that reflects native Mandarin phonological and phonetic influences. Imitation of Categorised tones with high percent choice and goodness ratings reflected strong native phonological influence and low sensitivity to phonetic differences between native and non-native phones. T241 was Categorised as M51 and T45 was Categorised as M35 (**Table 1**) with percent choice and goodness ratings in the high range. Imitations of T241 showed larger $F0_{excursion}$ and earlier $F0_{maxloc}$ than the target stimuli, consistent with the characteristics of M51 relative to T241. Similarly, T45 was imitated with lower $F0_{mean}$ and larger $F0_{excursion}$ than the original stimuli, consistent with the characteristics of M35, to which it had been assimilated. These deviations indicated that the phonetic details of imitations of both tones by Mandarin participants were constrained by L1 phonological features.

Conversely, imitation of non-native tones with moderate or low percent choice and goodness ratings was phonetically more accurate, suggesting high sensitivity to phonetic differences between native and non-native phones. T21 had been Categorised into M51 but with medium range of percent choice and goodness rating. T21 was imitated accurately with low deviation scores and was unaffected by the native category it had been assimilated to, M51, which has a much larger $F0_{excursion}$ and later $F0_{maxloc}$ than the stimulus.

Second, imitation was phonetically more accurate and less susceptible to native phonological influence in constant than variable talker blocks. Extending principles of ASP to imitation, imitators should use a more phonetic mode in constant talker blocks because within-talker phonetic information (e.g., in duration, $F0_{mean}$, and $F0_{excursion}$) is less variable and more reliable and thus is more easily detected. As a result, imitations should be more accurate in constant than variable talker blocks. Indeed, syllable duration for T21 and T45, $F0_{mean}$ for T241 imitations were more accurate in constant than in variable talker blocks as expected. Conversely, imitators should use a more phonological mode of perception in variable talker blocks in which phonetic information is variable and unreliable and thus they have to use their native phonological perceptual routines based on perceptual abstractions of tone contours and heights. Therefore, they should be more constrained by native language phonology. We reason that lower $F0_{mean}$ for T241 in the variable talker block reflects increased native phonological influence because T241 had been Categorised as M51 by the same participants, which has a lower $F0_{mean}$ than T241.

Finally, memory load interacted with tone for $F0_{mean}$, $F0_{excursion}$, and $F0_{maxloc}$. However, none of variations in $F0_{excursion}$ and $F0_{maxloc}$ deviation scores between the two memory load conditions were significant for the same tone. Only T315 had smaller deviations and lower $F0_{mean}$ values under low than high memory load, as we had predicted. This difference suggests that imitators showed high phonetic sensitivity to the T315 stimuli under low memory load, when phonetic details should be available in short memory, and which we expected to bias them to use a more phonetic mode of perception and imitation. On the other hand, they were expected to use a more phonological mode of perception under high memory load when the phonetic details had faded, and imitation should have become less accurate. Because T315 is an $UnCategorised_{clustered}$ tone assimilation with weak native phonological influences, however, this deviation cannot be attributed simply to native language phonological factors, but rather reflects sensitivities to non-contrastive phonetic details. Given that native Mandarin listeners are very sensitive to pitch direction (Gandour, 1983), we posit that the higher F0 value of a rising tone under high memory load may reflect that the higher F0 at the end of the Thai tone was better retained in memory than the lower F0 at the onset of the tone, due to recency effects. Consequently, under high memory load imitators tend to start their imitations of rising tones at a higher F0, resulting in an overall higher $F0_{mean}$ relative to the low memory load condition. This hypothesis could also explain why the $F0_{mean}$ of T45 imitation was unexpectedly more accurate but also higher in F0 value under high than low memory load. If imitators had activated a phonological mode of perception, T45 should instead have been affected by the native tone it was assimilated to (i.e., M35 which has a *lower* $F0_{mean}$ than T45). But they imitated with a higher rather than a lower F0. That is, we reason that higher $F0_{mean}$ in imitation of T45 under high memory load also cannot be attributed to native language phonological influence. We suggest it reflects instead a more phonetic-level tendency to start at a higher F0 when the phonetic details of the lower F0 at the onset of the tone have faded away while the higher F0 of the offset of the tone is still retained.

Our Mandarin experiment supported our hypotheses that both native phonological and phonetic factors, as reflected in assimilation patterns (from Chen et al., 2023), affect non-native imitation, and that their effect is modulated by memory load and talker variability, though in a restricted manner. In order to examine the generalizability of our findings to other non-Thai tone languages with differing tone systems, we tested imitation of the same Thai tones by Vietnamese native speakers in Experiment 2.

4. Experiment 2: Imitation of Thai tones by Vietnamese speakers

Vietnamese differs from Mandarin in its tone system, including the number of tones. Vietnamese listeners, accordingly, perceptually assimilated the same Thai tones into their native tone categories differently than Mandarin listeners, as indicated in a previous study (Chen et al., 2023). In Experiment 2, Vietnamese participants imitated Thai tones under the same manipulations of

memory load and talker/vowel variability. Their imitations were analysed with reference to their assimilation patterns to test the extension of our PAM- and ASP-driven hypotheses to Vietnamese.

4.1. Participants

Native speakers of Southern Vietnamese (n = 32), who were also living in Sydney to attend university or work, participated in Experiment 2, and were divided into two groups for the two imitation conditions (low memory load: M_{age} = 24.4 years, *SD* = 7.7 years, 12 females; high memory load: M_{age} = 27.2 years, *SD* = 12.8 years, 13 females). As with the Mandarin participants, all had first participated in a related study on perception of Thai tones (Chen et al., 2023) on the same day, and completed a background questionnaire before the test. Twenty-seven participants were born and raised in various locations in southern Vietnam (Ho Chi Minh City, Gia Lai, Phu Yen, Can Tho, Vinh Long, Da Nang, Bac Lieu, Tay Ninh, Binh Dinh, Ba Ria). The average length of residence in Australia for this group was 2.9 years (SD = 6.4 years, with three missing data points of length of residence). The remaining five, two for the low and three for the high memory load condition, were born to Vietnamese families in Australia. All 32 participants had acquired Southern Vietnamese as their native language and learned L2-English at school. All self-reported to have normal hearing and none had ever lived in Thailand or learned Thai or had more than two years of formal musical training. Stimulus materials, procedure, and data analysis are the same as Experiment 1.

4.2. Predictions

In a previous experiment (see Chen et al., 2023), the same participants had perceptually assimilated the same Thai tone stimuli used here (see **Table 3**). As in Experiment 1, we used those perceptual data as the basis for predictions about native phonological and phonetic influences on imitations of Thai tones. To quantify residual phonetic sensitivity for Vietnamese imitators, we first divided percent choice of the native tones above chance (20%, given their five-tone system) into three ranges: Low spanned 20%–46%, Medium 47%–74%, and High 75%–100%. For the category-goodness ratings, the scale was divided into the same three ranges as for the Mandarin participants: Low = 1.0–2.9, Medium = 3.0–4.9, and High = 5.0–7.0.

Under both memory loads (Chen et al., 2023), T21 was Categorised as V22; T241 was Categorised as V44; T315 was Categorised as V214 (see **Table 3**). For all three tones, percent choices were in the high range and goodness ratings varied from medium to high range, suggesting moderate residual phonetic sensitivity. T33 was $\text{UnCategorised}_{\text{clustered}}$ and assimilated to V44 and V22 under low memory load, suggesting a weak native phonological influence, but was Categorised as V22 under high memory load, suggesting a strong native phonological influence. In both cases, percent choices were in the medium range whereas ratings were in the high range, suggesting moderate residual phonetic sensitivity, which should moderately facilitate accurate imitation.

	Thai	T45		T33		T21		T315		T241	
	Response	**%**	**rating**	**%**	**rating**	**%**	**rating**	**%**	**rating**	**%**	**rating**
Low memory load	V44	6.3	3.5	**43.5**	5.4	4.3	3.6	–	–	**82.7***	5.7
	V22	2.2	2.4	**51.8**	5.3	**88.7***	5.4	1.8	3.4	13.2	5.4
	V35	**24.6**	4.7	1.1	2.2	–	–	7.6	5.8	1.8	3.5
	V21	**42.7**	4.3	3.6	3.6	6.3	3.9	5.6	3.5	1.6	3.8
	V214	**24.1**	4.3	–	–	–	–	**84.8***	5.2	–	–
Assimilation		$U_{clustered}$		$U_{clustered}$		C		C		C	
High memory load	V44	–	–	**38.1**	5.0	3.4	3.1	–	–	**81.1***	4.8
	V22	–	–	**61.6***	5.0	**93.7***	4.7	–	–	16.6	4.5
	V35	15.4	3.9	–	–	–	–	1.6	4.3	1.8	3.0
	V21	**60.9***	4.0	–	–	1.6	2.3	6.8	2.8	–	–
	V214	22.8	3.0	–	–	1.3	2.5	**91.4***	4.0	–	–
Assimilation		C		C		C		C		C	

Table 3: Assimilation of Thai tones into Vietnamese tone categories under low versus high memory loads (data from Chen et al., 2023). Categories in bold are choices that were significantly above chance: 20% for Vietnamese; "*" = Categorised tone. Assimilations: C = Categorised, U = UnCategorised. Ratings: 1 = Poor, 7 = Perfect; mean ratings are displayed. "–" = responses <1%.

T45 was also $UnCategorised_{clustered}$ and assimilated to V35, V21, V214 under low memory load with percent choices for these response categories in the low range and ratings in the medium range, suggesting a weak native phonological influence and moderate residual phonetic sensitivity to differences between the native and non-native tones. Imitation in this case should be phonetically accurate and less susceptible to native phonological constraints. But under high memory load, T45 was Categorised as V21 with percent choice and ratings in the medium range, suggesting a moderate native phonological influence, which should thus result in imitation moderately affected by the native phonological system.

4.3. Results

There were 31 missing data points (0.6%) for the Vietnamese group, as defined in Experiment 1; we excluded these data from the analysis. Four models were built to test all main effects and interactions, following the procedures described in Experiment 1. As in Experiment 1, initial models with participants as a random factor and random slopes for all within-subject factors converged, but were too complex to estimate *p*. Thus, we report here the main effects for multilevel factors or interactions of *F*-tests (see **Table 4**) with participants specified with

	Duration			**F0$_{mean}$**			**F0$_{excursion}$**			**F0$_{maxloc}$**		
	F	***df***	***p***	***F***	***df***	***p***	***F***	***df***	***p***	***F***	***df***	***p***
Tone	130.0	4, 5019	**<.001**	296.9	4, 5019	**<.001**	578.0	4, 5019	**<.001**	208.4	4, 5019	**<.001**
Talker (Tlk)	2.1	1, 5019	0.146	0.2	1, 5019	0.644	0.5	1, 5019	0.469	205.0	1, 5019	**<.001**
Vowel (V)	3.7	1, 5019	(0.054)	1.7	1, 5019	0.187	0.1	1, 5019	0.793	0.2	1, 5019	0.631
Memory (Mem)	0.0	1, 30	0.977	0.3	1, 30	0.580	0.7	1, 30	0.398	0.1	1, 30	0.781
Tone × Tlk	10.3	4, 5019	**<.001**	13.4	4, 5019	**<.001**	5.0	4, 5019	**0.001**	62.4	4, 5019	**<.001**
Tone × V	0.7	4, 5019	0.573	0.6	4, 5019	0.681	0.4	4, 5019	0.818	1.0	4, 5019	0.384
ToneÍMem	7.0	4, 5019	**<.001**	8.0	4, 5019	**<.001**	6.4	4, 5019	**<.001**	1.9	4, 5019	0.115
Tlk × V	0.2	1, 5019	0.679	0.4	1, 5019	0.533	1.5	1, 5019	0.215	0.5	1, 5019	0.468
Tlk ÍMem	1.0	1, 5019	0.319	1.7	1, 5019	0.192	0.5	1, 5019	0.499	0.4	1, 5019	0.545
VÍMem	0.7	1, 5019	0.412	5.7	1, 5019	**0.017**	0.0	1, 5019	0.844	0.0	1, 5019	0.912
TlkÍ V Í Mem	3.2	1, 5019	(0.072)	0.1	1, 5019	0.776	0.0	1, 5019	0.854	0.1	1, 5019	0.793
Tone × Tlk × V	0.2	4, 5019	0.931	0.5	4, 5019	0.717	0.3	4, 5019	0.860	1.9	4, 5019	0.104
Tone × TlkÍMem	0.6	4, 5019	0.637	1.7	4, 5019	0.140	1.1	4, 5019	0.335	0.0	4, 5019	0.999
Tone × VÍMem	0.3	4, 5019	0.868	0.7	4, 5019	0.603	1.1	4, 5019	0.378	0.4	4, 5019	0.804
Tone × Tlk × VÍMem	0.7	4, 5019	0.611	0.8	4, 5019	0.508	0.3	4, 5019	0.877	0.2	4, 5019	0.959

Table 4: Model details of acoustic measure deviation scores of Vietnamese imitators. Significant effects are in bold; marginal effects are in parentheses.

random intercept, as described in Experiment 1 (for deviation scores of imitations by Vietnamese participants, see Appendix A, Table A.3).

Main effects were significant for tone in all four acoustic-related deviation scores, and for talker variability in $F0_{maxloc}$. The main effects of memory load were non-significant, but there were memory load × tone interactions for syllable duration, $F0_{mean}$, and $F0_{excursion}$ and a memory load × vowel variability interaction for $F0_{mean}$.

To further examine the main effect of tone, we tested pairwise tone differences as in Experiment 1 (see Appendix C, Table C.1 for statistical details). For duration scores, all pairwise comparisons were significant except for that between T241 (M = 0.053, 95% CIs [0.047, 0.060]) and T33 (M = 0.061, 95% CIs [0.054, 0.067]). The confidence intervals included zero for T21 (M = –0.003, 95% CIs [–0.011, 0.005]) indicating a non-significant difference from the Thai target values (i.e., small deviations and hence the best imitation of duration). T315 (M = –0.036, 95% CIs [–0.044, –0.029]) and T45 (M = 0.018, 95% CIs [0.011, 0.024]) durations were imitated with moderate deviations.

For $F0_{mean}$, all pairwise comparisons were significant except for the comparison between T241 (M = –0.021, 95% CIs [–0.024, –0.017]) and T33 (M = –0.024, 95% CIs [–0.027, –0.021]). The confidence interval of the score included zero for T21 (M = –0.0004, 95% CIs [–0.003, 0.003]), indicating a non-significant difference form the Thai target, thus the best $F0_{mean}$ imitation. T315 imitation showed the largest positive score (M = 0.047, 95% CIs [0.042, 0.051]), indicating higher $F0_{mean}$ than the target stimuli. On the other hand, T33, T241, and T45 (M = –0.038, 95% CIs [–0.042, –0.033]) were all imitated with moderate negative scores, indicating lower $F0_{mean}$ than the stimuli.

Imitations of T315 (M = 0.244, 95% CIs [0.231, 0.256]), T33 (M = 0.095, 95% CIs [0.089, 0.101]), T45 (M = 0.056, 95% CIs [0.049, 0.064]), and T241 (M = 0.027, 95% CIs [0.019, 0.035]) had positive scores, indicating larger $F0_{excursion}$ than the stimuli. T315 showed the largest and thus least accurate $F0_{excursion}$, followed by T33, T45, and T241. The confidence interval of the score included zero for T21 (M = 0.001, 95% CIs [–0.005, 0.007]), indicating no difference from the Thai target stimulus value, and therefore the best $F0_{excursion}$ imitation.

For $F0_{maxloc}$, all the pairwise comparisons were significant. The confidence interval of the score included zero for T21 (M = 0.006, 95% CIs [–0.002, 0.015]), suggesting very accurate $F0_{maxloc}$ imitation. T315 imitation showed the only positive score (M = 0.082, 95% CIs [0.063, 0.101]), indicating later $F0_{maxloc}$ than the Thai target stimuli. T241(M = –0.143, 95% CIs [–0.154, –0.133]), T45 (M = –0.096, 95% CIs [–0.109, –0.083]) and T33 (M = –0.069, 95% CIs [–0.082, –0.056]) imitations had negative scores, indicating earlier $F0_{maxloc}$ than the Thai stimuli. T241 imitations showed the largest deviation (i.e., the least accuracy), followed by T45 and T33.

The main effect of talker variability indicates that imitation of $F0_{maxloc}$ was unexpectedly *less* accurate for constant (M = –0.084, 95% CIs [–0.090, –0.077]) than variable talker blocks

(M = –0.005, 95% CIs [–0.015, 0.006]), which should be interpreted with respect to tone, given the significant tone × talker variability interaction. Thus, we ran talker variability comparisons for each tone to examine the tone × talker variability interactions for all four acoustic measures (see Appendix C, Table C.3 for statistical details).

$F0_{maxloc}$ imitation was more accurate for T315 in constant talker blocks (M = –0.027, 95% CIs [–0.039, 0.016]) as the confidence interval of the score included zero (i.e., no difference from the Thai target stimuli) than it was in variable talker blocks (M = 0.190, 95% CIs [0.157, 0.224]). In contrast, that for T33 and T241 was less accurate in constant (M_{T33} = –0.133, 95% CIs [–0.149, –0.117]; M_{T241} = –0.179, 95% CIs [–0.192, –0.165]) than variable talker blocks (M_{T33} = –0.005, 95% CIs [–0.023, 0.013]; M_{T241} = –0.108, 95% CIs [–0.123, –0.093]). Earlier $F0_{maxloc}$ in T241 imitations renders the contour more like a falling tone (see **Figure 5**). Mandarin imitators also showed this pattern for T241. This was not expected. We speculate that because imitators tended to realise T241 with an earlier $F0_{maxloc}$ than the target, the higher talker variability in T241 than the other Thai stimulus tokens (see **Figure 3**) may have contributed to the greater mean deviations observed in constant talker blocks, in which imitators followed

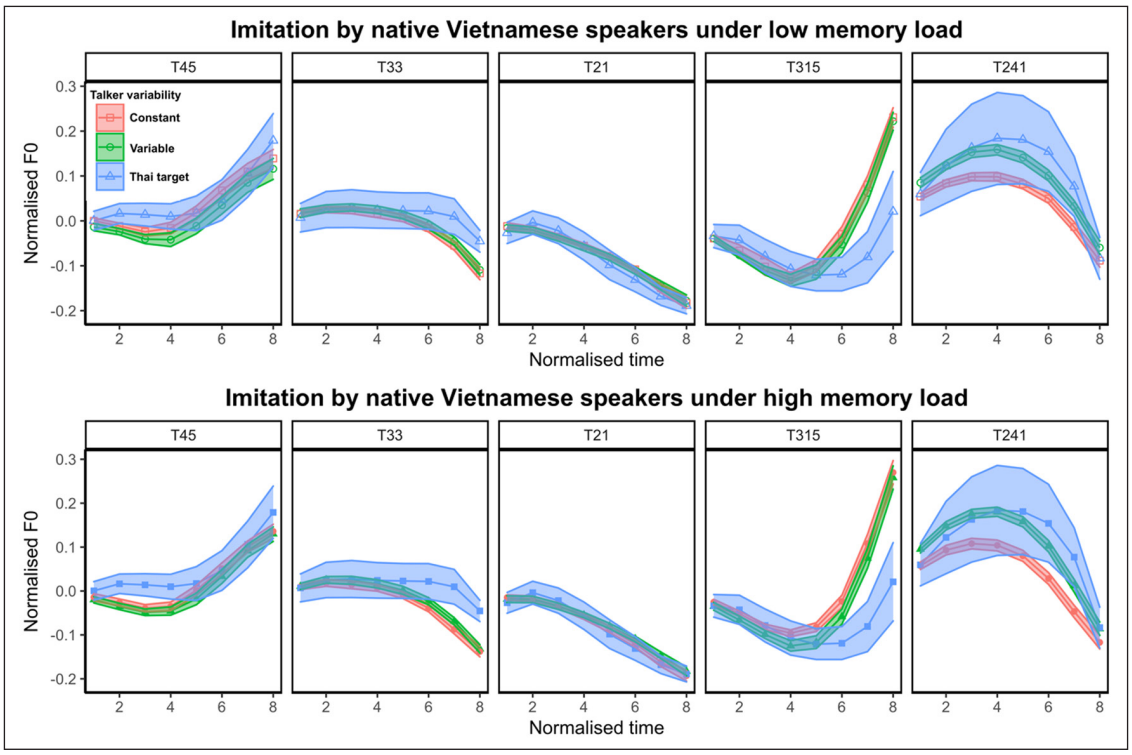

Figure 5: The time-and-Lobanov-normalised mean F0 contours of the Thai stimulus tones and their imitations by Vietnamese participants in the constant versus variable talker blocks. Ribbons indicate 95% confidence intervals for each mean contour type (blue = Thai stimuli; green = variable talker; red = constant talker).

the stimulus closely each trial but with earlier realisations of $F0_{maxloc}$. However, in variable talker blocks imitators appear to have used a more phonological mode of imitation of the tone type and realised T241 with similar $F0_{maxloc}$ independent of the $F0_{maxloc}$ of the specific stimulus token. When we calculated deviation scores for the variable talker block imitations, some of their $F0_{maxloc}$ realisations were actually similar to or even later than in the stimulus, resulting in a smaller mean deviation score.

T241 imitation of $F0_{mean}$ was more accurate in constant (M = –0.009, 95% CIs [–0.014, –0.004]) than variable talker blocks (M = –0.033, 95% CIs [–0.038, –0.027]). T21 duration was imitated longer in constant (M = 0.018, 95% CIs [0.007, 0.029]) and shorter in variable talker blocks (M = –0.024, 95% CIs [–0.034, –0.014]) as compared with the native Thai target.

Memory load effects are limited to interactions with tone and vowel variability. We broke down tone × memory load interactions for syllable duration, $F0_{mean}$, and $F0_{excursion}$ with memory load comparisons for each tone as in Experiment 1 (see Appendix C, Table C.2 for statistical details). Only T315 showed significant memory load differences. Both duration and $F0_{mean}$ indicated more accurate T315 imitation under low ($M_{duration}$ = –0.022, 95% CIs [–0.032, –0.013]; M_{F0mean} = 0.039, 95% CIs [0.033, 0.045]) than high memory load ($M_{duration}$ = –0.051, 95% CIs [–0.061, –0.040]; M_{F0mean} = 0.054, 95% CIs [0.048, 0.060]), as predicted.

In addition, we ran multiple pairwise comparisons with Tukey adjustments to tease apart the vowel variability × memory load interaction in $F0_{mean}$. Under high memory load, $F0_{mean}$ imitation was more accurate in constant (M = –0.005, 95% CIs [–0.009, –0.001]) than variable vowel blocks (M = –0.011, 95% CIs [–0.015, –0.008]), t = –2.621, df = 5019.1, p = 0.0436. No other comparisons were significant.

4.4. Discussion

First, imitations of Thai tones by Vietnamese participants were overall good with T21 being the best imitated tone. However, there existed tone-specific phonetic deviations that varied in strength. These variations reflected the unique native language influence as indicated by the participants' perceptual assimilation patterns. Categorised assimilation indicates strong native language influence and should constrain non-native imitation. T315 was Categorised with a high percent choice into V214, which has a larger $F0_{excursion}$ and later $F0_{maxloc}$ than T315 (see **Table 3**). T315 was imitated with larger $F0_{excursion}$ and later $F0_{maxloc}$, like V214. Similarly, T241 was Categorised as V44 with high percent choice, which is lower than T241 in $F0_{mean}$. T241 was imitated with lower $F0_{mean}$ than the target tone, and more like the native tone to which it had been perceptually assimilated, as we predicted. In both cases, high percent choice in categorisation indicates low residual sensitivity to phonetic differences between non-native and native tones, and leads to deviations that resemble the corresponding native tones.

On the other hand, for non-native tones that are Categorised with moderate percent choice and/or good ratings, listeners should display moderate residual phonetic sensitivity in perception and imitation. In the previous perception study (Chen et al., 2023; see **Table 3**), the same Vietnamese listeners as in the present study Categorised T21 into V22 with high percent choice, but the goodness rating was in the medium range under high memory load, suggesting moderate residual sensitivity to the differences between T21 and V22. V22 has shorter duration and smaller $F0_{excursion}$ than T21, but the imitation of T21 had the smallest deviation of all the Thai tones, showing little difference from the native Thai stimuli on all four measures. This suggests that native phonological influence was only moderate for T21 (i.e., Vietnamese imitators detected the phonetic details of the target stimuli and realised them relatively accurately in imitation).

Second, imitation was more accurate in constant than variable talker blocks, but this effect was limited to certain tones. Deviation scores of $F0_{maxloc}$ for T315, and of $F0_{mean}$ for T241 indicated more accurate imitation in constant than variable talker blocks. These results support our hypothesis that in constant talker blocks listeners were more sensitive to specific, concrete temporal and F0 properties of the stimuli, consistent with a more phonetic mode of perception, and consequently imitated the stimuli more accurately than in variable talker blocks, where they appear to have used a more phonological mode of perception in which native phonological perceptual routines were activated, constraining imitation. T315 was Categorised as V214, which has a larger $F0_{maxloc}$. Compatibly, imitation of $F0_{maxloc}$ for T315 was later, as indicated by positive deviations, in variable than constant talker blocks. Similarly, T241 was Categorised as V44, which has a lower $F0_{mean}$, and imitation of T241 had a lower $F0_{mean}$ in variable than constant blocks. In addition, T21 was Categorised as V22 which has a shorter syllable duration. Syllable duration of T21 in imitation was shorter in variable talker blocks but longer in constant talker blocks as compared with the native norm. These findings support our hypothesis that imitation should reflect a larger native phonological influence in variable than constant talker blocks.

Third, memory load showed no main effect but interacted with tone for syllable duration, $F0_{mean}$, and $F0_{excursion}$. However, as was observed in Experiment 1 for the Mandarin imitators, only imitation of T315 showed significant differences between the two memory load conditions. Imitation of T315 was more accurate in terms of syllable duration and $F0_{mean}$ under low memory load, when rich phonetic details of F0 properties are still available and a more phonetic mode of perception is activated, as we expected. In addition, when rich phonetic details have decayed and imitators use a more phonological mode of perception under high memory load, duration was shorter in imitations than in the Thai target stimuli. This reflects native phonological constraints, as the native category that T315 was assimilated to (i.e., V214) also has a shorter duration than T315.

Fourth, the effect of vowel variability on $F0_{mean}$ was significant only when the memory load was high, which is compatible with our predictions. Under high memory load, imitation of $F0_{mean}$ was more accurate in constant than variable vowel blocks. We argue that under high memory

load, listeners used a more phonological mode of perception, reducing their ability to process concrete, fine-grained F0 details of the stimuli. Consequently, imitation was less accurate in variable vowel blocks, where imitators have to abstract phonological features from the more variable stimulus, than constant vowel blocks.

5. General discussion

The two experiments reported here demonstrate that native phonological and phonetic influences as indicated in the perceptual assimilation results (see Chen et al., 2023) predicted the strength of phonetic variations in non-native tone imitation, which were also modulated by stimulus variability and memory load that shifted imitators' modes of perception/imitation for specific tones that differed between the two groups.

5.1. Effects of native language phonological and phonetic factors

The main effects of tone showed that phonetic accuracy of imitations varied in strength for all measures, although imitations of certain Thai tones, such as T21, were generally good. These phonetic deviations in the imitations reflect native language phonological constraints in perception as indicated by assimilation types, whereas phonetic imitation accuracy is commensurate with residual phonetic sensitivity in perception as indicated by percent choice and goodness ratings. For Categorised tones with high percent choice and/or goodness ratings, native phonological influences constrained non-native imitation, resulting in deviations similar to characteristics of the corresponding native tones to which they had been assimilated. For Mandarin imitators, deviations in their imitations of T241 and T45 were compatible with the native Mandarin tones that they were assimilated to. Similarly, for Vietnamese imitators, T315 was affected by the native category that it was assimilated to (i.e., V214).

Even when a non-native tone was phonologically Categorised as a native tone, we found that residual phonetic sensitivity to the differences between native and non-native tones facilitated imitation of non-native tones. T21 was Categorised into M51 by Mandarin listeners and into V22 by Vietnamese listeners (Chen et al., 2023), but listeners of both language groups showed moderate residual phonetic sensitivity to differences between these native and non-native tones. T21 was the best imitated tone with low deviation scores in all measures for both language groups and was not affected by the native tone it was assimilated to. This suggests that imitators retained phonetic details of the T21 target stimuli and instantiated them in their imitations.

There was only one type of UnCategorised assimilation (i.e., $\text{UnCategorised}_{\text{clustered}}$) in the present study. We posited that non-native tones of this type should bear weak native phonological effects as listeners did not perceive strong phonological similarity to any single native category but weak similarities to two or three native categories. Consequently, their imitation should not be affected by any single native category. For Mandarin imitators, T315 was split between

M35 and M214, and did not reach the categorisation threshold for any single category. Its deviations in imitation cannot be attributed to influence from either M35 or M214 alone. For Vietnamese imitators, assimilation of T45 and T33 were $\text{UnCategorised}_{\text{clustered}}$ under low memory load, reflecting weak phonological influence, but were Categorised under high memory load, indicating strong phonological influence. However, imitations of the two tones across both memory loads were not significantly different from the Thai target values for any deviation measures. Given that imitation requires phonetic level details, and percent choice and goodness ratings of native response categories under both memory load conditions were in the low-to-medium range and thus comparable, we speculate that the comparable phonetic sensitivity outweighed the differences in phonological influence indicated by difference assimilation types.

Mandarin and Vietnamese imitators have different native tone systems with different phonetic realisations of each native phonological categories. These differences clearly affected some aspects of their imitations as reflected in the comparison of the same tone imitated by the two language groups (see Table C.4 in Appendix C for full statistical details). For example, T21 was Categorised as M51 by Mandarin participants, which has a larger $\text{F0}_{\text{excursion}}$ than V22, to which the Vietnamese participants had assimilated T21. Although T21 was among best imitated tones for both Mandarin and Vietnamese groups, $\text{F0}_{\text{excursion}}$ was larger in Mandarin ($M = 0.053$, 95% CIs [0.045, 0.061]) than Vietnamese imitations ($M = 0.001$, 95% CIs [–0.005, 0.007]). Similarly, Mandarin participants had Categorised T241 to M51, which has a larger $\text{F0}_{\text{excursion}}$ than the Vietnamese high level tone V44 to which Vietnamese group had Categorised T241. Although T241 was imitated with larger $\text{F0}_{\text{excursion}}$ than the stimuli by both language groups, the difference was larger for Mandarin ($M = 0.102$, 95% CIs [0.089, 0.114]) than Vietnamese participants ($M = 0.027$, 95% CIs [–0.035, 0.019]). In these cases, imitations of Categorised non-native tones were affected by each group's native tone phonological features, supporting the extension of the PAM principle that Categorised assimilation imparts a strong native phonological influence on imitation performance.

5.1. Effects of talker and vowel variability

Talker and vowel variability showed some restricted main effects for each group. Imitation was more accurate in constant than variable talker blocks for both language groups. T21 and T45 imitation of syllable duration by Mandarin participants, T315 imitation of $\text{F0}_{\text{maxloc}}$ by Vietnamese participants, and T241 imitation of F0_{mean} by both groups were more accurate in constant than variable talker blocks. We argue that in constant talker blocks, where specific temporal and F0 properties are constant and reliable, imitators are biased to use a more phonetic mode of perception and focus more on phonetic level information. Consistent with this reasoning, their imitation was phonetically more accurate. On the other hand, in variable talker blocks, we posited that imitators would be biased to use a more phonological mode of perception because the low-

level phonetic information is variable, and they should therefore rely on abstract phonological temporal and pitch features via their native phonological perceptual routines. In non-native imitation, the phonological mode should result in imitation that is less phonetically accurate and more constrained by native phonological features. This hypothesis is supported by our findings that imitation in variable talker blocks displayed features such as $F0_{mean}$ and $F0_{maxloc}$, which were constrained by native tones.

Vowel variability appears to be easier to process than talker variability, as it affected only a small number of measures and only for Vietnamese participants. Under high memory load, the Vietnamese group imitated $F0_{mean}$ more accurately in constant than variable vowel blocks. For Mandarin imitators, vowel variability did not yield any main effects or interactions for any of the deviation scores we examined.

In a previous perception study (Chen et al., 2023), we had found no significant effects of talker or vowel variability in categorisation by either the Mandarin or the Vietnamese group, but discrimination was more accurate in constant talker/vowel blocks than variable talker/vowel blocks for both language groups. We reasoned that discrimination requires listeners to compare two non-native phones phonetically, which can be best accomplished via phonetic level processing and thus listeners focused more on details at the phonetic level. In this sense, imitation may be more similar to discrimination than to categorisation in that imitators need to reproduce the phonetic details of the non-native tone and consequently must attend to the phonetic details of the target stimuli, which is somewhat hindered by talker/vowel variability because it biases toward a more phonological mode of perception.

5.2. Effects of memory load

According to the principles of ASP, under low memory load, listeners retain rich phonetic details of F0 properties in working memory and thus are biased toward using a more phonetic mode of perception. In this mode, listeners attend more to phonetic details and are less constrained by native language phonological pitch features. On the other hand, under high memory load, phonetic details decay in working memory, and a more phonological mode of perception should be activated. In the phonological mode, listeners are less sensitive to phonetic details because they have faded, thus they have only more lasting perceptual abstractions available in memory. Consequently, their native language phonological system has more influence. Extending these principles to imitation, it should be phonetically more accurate under low than high memory load.

Our study on tone imitation did not find main effects of memory load, consistent with previous studies on vowel (Repp & Williams, 1985) and segmental length (Asano & Braum, 2016) imitations. However, we did find some evidence of memory load effects as modulated by tone, mostly for T315. Indeed, for Mandarin imitators, $F0_{mean}$ for T315 was imitated more

accurately under low than high memory load. For Vietnamese imitators, similarly, imitation of syllable duration and $F0_{mean}$ for T315 was more accurate under low than high memory load. These results are consistent with our hypothesis that listeners should be biased toward a more phonetic mode of perception under low memory load, when phonetic details in short-term memory are still available, and thus to imitate non-native F0 properties more accurately. Under high memory load, however, we reason that the rich array of fine-grained phonetic details in short-term memory will have faded, shifting imitators toward a more phonological mode of perception. In this mode, native phonological perceptual routines should be more activated, biasing listeners to imitate Thai tones with abstract phonological features of native tones. In line with this, Vietnamese listeners Categorised T315 as V214, which is shorter than T315, and their imitations of T315 were shorter under high than low memory load, consistent with a phonological influence from native tone V214. The fact that memory load effects were limited mostly to a phonetically complex falling-rising tone suggests that the phonetic details of simple tones, such as phonologically-defined level, rising or falling tones, are less susceptible to decay in memory and thus easy to imitate.

5.3. Limitations and future directions

We must note limitations of the present study. First, we examined non-native tone imitation only by native tone language listeners with no experience with the unfamiliar language, Thai. We did not examine L2 learners of Thai from these L1 groups. We can cautiously suggest that accurate imitation of some tones of a new language are likely to be facilitated for L2 learners who have a tone language as an L1, specifically in relation to their perceptual assimilation of specific L2 tones to their native tone systems. But this would need to be investigated directly with L2 tone language learners of native tone languages. Importantly, the present study suggests that neither Mandarin nor Vietnamese learners of L2-Thai would have an absolute advantage for imitating all tones. They should, instead, display PAM-consistent differences in which tones would show better/worse L2 tone imitations by each group of learners, in line with their assimilations of the individual tones to their native tone systems. To fully understand L2 development in terms of lexical tones, future research should investigate imitation/production of the same Thai tones by Mandarin and/or Vietnamese learners of Thai with differing L2-Thai proficiency.

Second, our manipulation of memory load may not have been sufficient to test our hypothesis about memory demands leading to shifts between phonological and phonetic processing of the target stimuli. Future research should use longer intervals and add a secondary task before imitation, such as counting digits, to induce a high memory load, which could substantially increase the effect of memory load on imitation. Another way to vary memory load would be to

elicit a series of imitations, with the first repetition representing the condition of low memory load and the last one representing high memory load (Cole & Shattuck-Hufnagel, 2011).[5] It is expected that participants will activate a more phonological mode of imitation when memory load is high enough that short-term memory of phonetic details has more completely faded and is no longer available. Under such conditions, we expect imitation to be more constrained by native phonology than we were able to observe under the current high memory load condition, a delay of 2000 ms without an intervening task.

Moreover, a native Thai control group could be included for a better interpretation of the memory load effect. Native imitators should be less susceptible to memory load manipulation than non-native imitators as the task should be akin to a repetition task for native listeners, i.e., more focused on phonological categories than phonetic details and thus less susceptible to memory load manipulations. In addition, future research could also explore the role of the individual variability in coping with memory loads by measuring their working memory capacities as in some recent studies (e.g., Petrone, D'Alessandro, & Falk, 2021).

6. Conclusion

In conclusion, deviations in imitation varied among different Thai tones and reflected native phonological influences as indicated by the participant groups' perceptual assimilation patterns, in line with PAM-based predictions and supporting the articulatory commonality between non-native perception and imitation assumed by PAM. Although non-native listeners Categorised non-native tones into their native categories, they also could retain moderate residual sensitivity to phonetic details, indicated by percent choice and goodness ratings. When imitating non-native tones, they used this residual sensitivity to produce phonetically more accurate imitations.

Memory load and stimulus variability moderately affected accuracy in non-native imitation of lexical tones. Imitations of certain tones were generally less accurate and more constrained by native language phonological features under high memory load and in variable talker/vowel blocks where imitators presumably used a more phonological mode of perception and had lower sensitivity to concrete temporal and F0 properties. On the other hand, under low memory load and in constant talker/vowel blocks, imitators used a more phonetic mode of perception in which they showed more sensitivity to concrete temporal and F0 properties and produced phonetically more accurate imitations.

[5] Note that imitators' motivation can affect imitation as suggested by an anonymous reviewer. That is, after a series of practice or test trials, some imitators may learn to imitate the model more closely. However, we argue that such individual variations could be accounted for at the statistical modelling stage.

The current findings thus have substantive implications for theories about the effects of language experience on speech perception and production. Phonetic accuracy in non-native imitation is commensurate with the amount of residual phonetic sensitivity in perception. Deviations from the target stimuli in non-native imitation can be at least partially traced to the imitators' native phonological constraints, which are predictable from their perceptual assimilation patterns. Native language phonological constraints and residual phonetic sensitivities as well as phonological and phonetic modes of perception should be considered when researching non-native tone imitation and learning. On the applied side, the results suggest that teachers of tone languages should tailor their pedagogy to address potential problems caused by native phonological influences for students of different language backgrounds, particularly those who speak other tone languages.

Additional file

The additional file for this article can be found as follows:

- **Supplementary file.** Appendix A to D. DOI: https://doi.org/10.5334/labphon.6435.s1

Funding information

This work was supported by the China Foreign Language Education Fund [grant number ZGWYJYJJ11Z007] and a China Scholarship Council and Western Sydney University Joint PhD study scholarship awarded to the first author.

Competing interests

The authors have no competing interests to declare.

References

Alivuotila, L., Hakokari, J., Savela, J., Happonen, R.-P., & Aaltonen, O. (2007). Perception and imitation of finnish open vowels among children, naïve adults, and trained phoneticians. *Proceedings of ICPhS 2007*, 361–364.

Asano, Y. (2018). Discriminating non-native segmental length contrasts under increased task demands. *Language and Speech*, *61*(3), 409–429. DOI: https://doi.org/10.1177/0023830917731907

Asano, Y., & Braum, B. (2016). Does speech production in L2 require access to phonological representations? *Proceedings of the International Conference on Speech Prosody*, 237–241. DOI: https://doi.org/10.21437/SpeechProsody.2016-49

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. DOI: https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. DOI: https://doi.org/10.1016/S0079-7421(08)60452-1

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

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

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

Best, C. (2015). Devil or angel in the details?: Perceiving phonetic variation as information about phonological structure. In J. Romero & M. Riera (Eds.), *Phonetics-phonology interface: Representations and methodologies* (pp. 3–31). John Benjamins Publishing Company. DOI: https://doi.org/10.1075/cilt.335.01bes

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*. DOI: https://doi.org/10.3389/fpsyg.2019.00364

Best, C., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, *20*(5), 539–542. DOI: https://doi.org/10.1111/j.1467-9280.2009.02327.x

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 13–34). John Benjamins Publishing Company. DOI: https://doi.org/10.1075/lllt.17.07bes

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, *37*(1), 79–96. DOI: https://doi.org/10.1016/j.wocn.2008.09.003

Burnham, D., Kuratate, T., McBride-Chang, C., & Mattock, K. (2009). *Making speech three-dimensional: Adding tone to consonant- and vowel-based speech perception and language acquisition research, quantification and theory*. http://purl.org/au-research/grants/arc/DP0988201

Carignan, C. (2018). Using naïve listener imitations of native speaker productions to investigate mechanisms of listener-based sound change. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), Article 1. DOI: https://doi.org/10.5334/labphon.136

Chao, Y. R. (1930). A system of tone-letters. *Le Maitre Phonetique*, *45*, 24–27.

Chen, J., Antoniou, M., & Best, C. (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. DOI: https://doi.org/10.1016/j.wocn.2022.101199

Chen, J., Best, C., & Antoniou, M. (2020). Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai tones by Mandarin and Vietnamese listeners. *Journal of Phonetics*, *83*, 101013. DOI: https://doi.org/10.1016/j.wocn.2020.101013

Chen, J., Best, C., Antoniou, M., & Kasisopa, B. (2018). *Mapping and comparing East and Southeast Asian language tones*. Australia Linguistic Society annual conference, Adelaide.

Chen, J., Best, C., Antoniou, M., & Kasisopa, B. (2019). Cognitive factors in perception of Thai tones by naïve Mandarin listeners. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1684–1688). Australasian Speech Science and Technology Association Inc.

Cole, J., & Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: What do listeners imitate? *Proceedings of the Annual Conference of the International Speech Communication Association*, 969–972. DOI: https://doi.org/10.21437/Interspeech.2011-395

Faris, M. M., Best, C., & Tyler, M. D. (2018). Discrimination of uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics*, *70*, 1–19. DOI: https://doi.org/10.1016/j.wocn.2018.05.003

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229–273). York Press.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and emprical progress* (pp. 3–83). Cambridge University Press. DOI: https://doi.org/10.1017/9781108886901.002

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America*, *83*(2), 729–740. DOI: https://doi.org/10.1121/1.396115

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467–491. DOI: https://doi.org/10.1016/S0167-6393(02)00128-0

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*(1), 3–28. DOI: https://doi.org/10.1016/S0095-4470(19)30607-2

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gandour, J. (1983). Tone perception in far eastern languages. *Journal of Phonetics*, *11*(2), 149–175. DOI: https://doi.org/10.1016/S0095-4470(19)30813-7

Gandour, J. T. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 41–76). Academic Press. DOI: https://doi.org/10.1016/B978-0-12-267350-4.50007-8

Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *The Journal of the Acoustical Society of America*, *115*(5), 2545–2545. DOI: https://doi.org/10.1121/1.4783674

Halekoh, U., & Hojsgaard, S. (2014). A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models–the R package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–30. DOI: https://doi.org/10.18637/jss.v059.i09

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269–279. DOI: https://doi.org/10.1016/j.wocn.2011.11.001

Hao, Y.-C., & de Jong, K. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics*, *54*, 151–168. DOI: https://doi.org/10.1016/j.wocn.2015.10.003

Jia, G., Strange, W., Wu, Y., Collado, J., & 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*(2), 1118–1130. DOI: https://doi.org/10.1121/1.2151806

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*(2B), 606–608. DOI: https://doi.org/10.1121/1.1912396

Nguyen, N., Dufour, S., & Brunellière, A. (2012). Does Imitation Facilitate Word Recognition in a Non-Native Regional Accent? *Frontiers in Psychology*, *3*. DOI: https://doi.org/10.3389/fpsyg.2012.00480

Nhàn, N. T. (1984). *The syllabeme and patterns of word formation in Vietnamese* [PhD Dessertation]. New York University.

Nusbaum, H., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka & E. Vatikiotis-Bateson (Eds.), *Speech perception, production and linguistic structure* (pp. 113–134). Ohmsha Publishing. https://www.academia.edu/283545/Paying_Attention_to_Differences_Among_Talkers

Petrone, C., D'Alessandro, D., & Falk, S. (2021). Working memory differences in prosodic imitation. *Journal of Phonetics*, *89*, 101100. DOI: https://doi.org/10.1016/j.wocn.2021.101100

R Core Team. (2018). *R: A language and environment for statistical computing* [Manual]. https://www.R-project.org/

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention Perception & Psychophysics*, *77*(2), 571–591. DOI: https://doi.org/10.3758/s13414-014-0791-3

Repp, B. H., & Williams, D. R. (1985). Categorical trends in vowel imitation: Preliminary observations from a replication experiment. *Speech Communication*, *4*(1–3), 105–120. Scopus. DOI: https://doi.org/10.1016/0167-6393(85)90039-1

Rojczyk, A. (2012). Phonetic and phonological mode in second-language speech: VOT imitation. *EUROSLA 2012*, Poznań Poland.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*(3), 422–429. DOI: https://doi.org/10.3758/BF03194890

So, C. K., & Best, C. (2010a). Discrimination and categorization of Mandarin tones by Cantonese speakers: The role of native phonological and phonetic properties. *Proceedings of the 13th Australasian International Conference on Speech Science and Technology*.

So, C. K., & Best, C. (2010b). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293. DOI: https://doi.org/10.1177/0023830909357156

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. DOI: https://doi.org/10.1016/j.wocn.2010.09.001

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033–1043. DOI: https://doi.org/10.1121/1.1531176

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44. DOI: https://doi.org/10.3758/BF03207136

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, *75*(6), 1866–1878. DOI: https://doi.org/10.1121/1.390988

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter-and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*(2), 413–421. DOI: https://doi.org/10.1044/1092-4388(2003/034)

Xu, Y. (2013). *ProsodyPro—A tool for large-scale systematic prosody analysis*. TRASP 2013.

Yip, M. (2002). *Tone*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139164559