



Research Article

Phonological and phonetic contributions to perception of non-native lexical tones by tone language listeners: Effects of memory load and stimulus variability

Juqiang Chen^{a,*}, Mark Antoniou^b, Catherine T. Best^{b,c,*}

^a School of Foreign Languages, Shanghai Jiao Tong University, Shanghai 200240, China

^b Western Sydney University, The MARCS Institute for Brain Behaviour and Development, Penrith, NSW 2751, Australia

^c Haskins Laboratories, New Haven, CT, USA

ARTICLE INFO

Article history:

Received 21 February 2022

Received in revised form 1 November 2022

Accepted 14 November 2022

Keywords:

Non-native lexical tone perception

Perceptual assimilation

Talker variability

Vowel variability

Memory load

ABSTRACT

The present study examined native language phonological and phonetic factors in non-native lexical tone perception by tone language listeners, manipulating memory load and stimulus variability to bias listeners towards a more phonological or more phonetic mode of perception. Mandarin and Vietnamese listeners categorised the five Thai lexical tones to their native tones, and discriminated five selected Thai tone contrasts that were predicted by the Perceptual Assimilation Model (PAM, Best, 1995) to be discriminated differently. Categorisation responses showed more phonologically-based patterns under high than low memory load but were unaffected by talker and vowel variability, whereas discrimination accuracy was reduced by talker and vowel variability but not by memory load. Phonological factors indicated by type of categorisation and category overlap generally predicted the discrimination of non-native tone contrasts in line with PAM principles. Phonetic factors reflected in category overlap scores and fit index difference scores predicted variations in discriminating contrasts of the same contrast categorisation type. These findings uphold the extension of PAM principles to non-native tone perception by native listeners of other tone languages. Native phonological and phonetic contributions to non-native speech perception differ between categorisation and discrimination tasks, as reflected in differential modulation by memory load and stimulus variability.

© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

Perception of non-native phones is susceptible to native language influences at both phonological and phonetic levels. If a phonological contrast does not exist in the native language, listeners may have difficulty in identifying and discriminating it, e.g., native Japanese speakers have difficulty discriminating the English /r/-/l/ contrast, which is absent in Japanese (MacKain, Best, & Strange, 1981; Miyawaki et al., 1975; Sheldon & Strange, 1982; Yamada & Tohkura, 1992). Even when a contrast in an unfamiliar language exists in the listener's native language, differences in phonetic realisation

can also affect perception (Best & Strange, 1992; Hallé, Best, & Levitt, 1999). For example, /w/-/r/ is phonologically contrastive in both English and French but /r/ in French is a uvular fricative, approximant or trill and thus differs phonetically from the English coronal approximant /r/. Consequently, when French listeners were presented with an English /w/-/r/ continuum, they categorised /r/ less consistently than English listeners. Moreover, these phonetic effects can be substantially modulated by memory load (Asano, 2018; Mitterer & Mattys, 2017; Pisoni, 1973; Werker & Logan, 1985; Werker & Tees, 1984; Yu, Shafer, & Sussman, 2017), talker variability (Magnuson & Nusbaum, 2007; Mullennix & Pisoni, 1990), and phonetic context variability (Shaw & Tyler, 2020; Zheng, 2014), which can shift listeners to focus less on phonetic details and rely on a more abstract phonological mode of perception.

Most cross-language research has focused on perception of non-native consonants and vowels. Fewer studies have

* Corresponding authors at: Western Sydney University, The MARCS Institute for Brain Behaviour and Development, Penrith, NSW 2751, Australia (C.T. Best); Shanghai Jiao Tong University, School of Foreign Languages, No. 800 Dongchuan Road, Minhang District, Shanghai, P. R. China, 200240 (J. Chen).

E-mail addresses: Juqiang.c@sjtu.edu.cn (J. Chen), C.Best@westernsydney.edu.au (C.T. Best).

investigated how these factors affect non-native perception of lexical tones, particularly by listeners of other tone languages. One such study found both native phonological and phonetic effects on perceptual assimilation of Thai tones by native listeners of Mandarin and Vietnamese, both of which use lexical tones but differ in their tone systems (Chen, Best, & Antoniou, 2020). However, that study did not examine discrimination. The present study made use of those assimilation results to predict how Mandarin and Vietnamese listeners would discriminate different types of Thai tone contrasts, and also aimed to reconfirm the assimilation patterns. In the next section, speech perception theories will be reviewed regarding how they account for native language phonological and phonetic effects.

1.1. Native language phonological and phonetic effects on non-native perception

Several theoretical frameworks have been proposed to account for the influences of listeners' native language on non-native speech perception. As the focus of this paper is on the perception of non-native tone *contrasts*, the Speech Learning Model (SLM, Flege, 1995) and the Native Language Magnet model (NLM, Kuhl & Iverson, 1995), which focus more on individual phonetic categories rather than on phonological contrasts, will not be considered further. The Second Language Linguistic Perception model (L2LP: e.g., Escudero, 2005; Escudero & Vasiliev, 2011; Escudero & Williams, 2011) will also not be considered further here, as it has focused mainly on predictions based on acoustic similarities, and on perception of vowels rather than on tones, and more importantly it does not address phonological versus phonetic contributions nor does it address how these two factors are modulated by cognitive/variability manipulations which are central to the present study. The Perceptual Assimilation Model (PAM, Best, 1995) was selected as the theoretical framework for the present study because it provides the most coherent account for our purposes of examining how native phonological and phonetic properties affect perceptual assimilation of non-native phones into native categories, and how varying assimilation patterns influence discrimination of non-native contrasts.

PAM (Best, 1995) posits that naïve adult listeners categorise (perceptually assimilate) non-native phones to native phonemes based on the perceived articulatory gestural (articulatory-phonetic) similarities between them. A non-native phone can be heard with respect to the native phonological system (1) as an excellent, good, fair, mediocre or poor exemplar of a native phoneme, i.e., Categorized, or (2) as not falling within any single native phoneme but still falling within the native phonological space, i.e., Uncategorised or (3) as a non-speech sound, i.e., Non-Assimilable.

When the two phones of a non-native contrast are Categorized into contrasting native categories, i.e., Two-Category assimilation, they are predicted to be better discriminated than when two non-native phones are Categorized into the same native category but differ in their degrees of perceived phonetic discrepancy from the native "ideal", i.e., Category-Goodness assimilation. When two non-native phones are Categorized into the same native category as equally good or bad exem-

plars of that category, i.e., Single-Category assimilation, these contrasts will be the most difficult to discriminate.

Uncategorised-Categorised pairs should be discriminated nearly as well as Two-Category assimilations, because they involve a distinction across a native phonological boundary, i.e., falling within a native category versus outside of that category. Discrimination of Uncategorised-Uncategorised assimilations, however, should range from being poor to very good depending on the degree of phonetic differentiation and phonological category overlap between the two Uncategorised non-native phones.

Perceived phonological overlap between contrasting non-native phones can, in fact, decrease discrimination accuracy even within Two-Category, Uncategorised-Categorised and Uncategorised-Uncategorised assimilations (Antoniou, Best, & Tyler, 2013; Best, Avesani, Tyler, & Vayra, 2019; Faris, Best, & Tyler, 2018; So & Best, 2014). *Non-overlap* contrast assimilations refer to the cases where the contrasting non-native phones are each identified with completely different sets of native categories. When there are one or more shared categories for assimilation of the contrasting non-native phones, but not all choices are shared, this is designated *partial-overlap* assimilation. Finally, when all the response categories are shared for both non-native contrasts, this is considered a *complete-overlap* assimilation. Within each contrast assimilation type above, *non-overlap* contrasts are facilitated by native phonological distinctions and should be better discriminated than *partial-overlap* contrasts that are only partially assisted by native distinctions, which in turn should be better discriminated than *complete-overlap* contrasts that are not aided at all by native distinctions. Native phonological similarity interferes less in partial-overlap than in complete-overlap cases (Best et al., 2019).

Relatively few studies have investigated non-native tone perception, and they often examined only categorisation of synthetic tone continuum (Burnham & Jones, 2002) or only categorisation of non-native tones as the non-native categories, also called identification (Chang, Yao, & Huang, 2017; Wang, 2013), or only categorisation of non-native tones as native categories, also called assimilation (Chen et al., 2020; Wu, Munro, & Wang, 2014) or only discrimination (Burnham et al., 1996; Hao, 2017; Lee, Vakoch, & Wurm, 1996; Tsukada & Kondo, 2018; Wayland & Guion, 2004). Note that we use categorisation as an umbrella term for both identification and assimilation, which contrasts with discrimination. Of the few that examined both categorisation and discrimination of non-native tones (Reid et al., 2015; So & Best, 2014; Wu, Bundgaard-Nielsen, Baker, Best, & Fletcher, 2015), some tested only native listeners of non-tone languages (So & Best, 2014). This type of listener is likely to categorise non-native tones into their native intonation/prosodic categories rather than segmental categories, which is likely to differ qualitatively from assimilation to corresponding native tone categories by listeners of other tone languages (Best, 2019), in which tones serve segmental-level functions (Duanmu, 1990, 1994; Lin, 1989). Thus, the relationship between perceptual assimilation and discrimination of non-native tone contrasts could be different for listeners of other tone languages than for non-tone language listeners. Although some other studies have tested both assimilation and discrimination of non-native tones by native

tone language listeners, e.g., Mandarin listeners perceiving Thai tones (Reid et al., 2015) or Cantonese tones (Wu et al., 2015), they only tested one native tone language group and thus were unable to compare native language phonological and phonetic constraints across different tone systems. Consequently, it remains unresolved how native phonological and phonetic factors contribute to non-native lexical tone perception.

1.2. *Phonetic versus phonological mode of speech perception*

Performance in online processing of non-native speech by adult naïve listeners and second language learners varies across tasks and conditions. The Automatic Selective Perception model (ASP, Strange, 2011) proposes that listeners can switch between, or vary the balance between, phonological and phonetic modes of perceiving native and non-native speech. When the attentional focus of a task is on distinguishing phonetic differences that are essential to lexical distinctions, i.e., phonological distinctiveness (Best, 2015; Best, Tyler, Gooding, Orlando, & Quann, 2009), the phonological mode of speech perception is maximised. On the other hand, when the task requires listeners to attend to finer-grained phonetic details that are irrelevant to lexical distinctions, the phonetic mode is maximised to detect within-category phonetic variations.

In non-native perception, selective perception routines that are attuned to the native phonological system are automatically used in the phonological mode, leading to the strongest native phonological influence. Consequently, phonetic differences within a native phonological category are less likely to be detected. However, in the phonetic mode, non-native phonetic differences that fall within a single native phonological category are more likely to be detected than when they are processed in a phonological mode.

Most studies that have examined different modes of perception in non-native speech perception so far have only used discrimination but not categorisation tasks (Asano, 2018; Werker & Tees, 1984; Werker & Logan, 1985). Discrimination emphasises the detection of phonetic distinctions whereas categorisation involves attending to higher order invariant information in order to select the phonological category label. Thus, categorisation inherently requires a phonological level of processing whereas discrimination is more likely to rely on processing at the phonetic level. It remains unresolved how listeners alternate, or balance, between phonological and phonetic modes in discrimination and categorisation as a function of other factors, such as memory load and stimulus variability that shift processing toward or away from concrete phonetic details. In the next two sections, we synthesise existing findings of memory load and stimulus variation effects on speech perception and provide our predictions and hypotheses for the effects of these manipulations on both categorisation and discrimination tasks in light of PAM and ALS.

1.2.1. *Memory load*

The availability of phonetic details in short-term memory determines listeners' ability to use a phonetic mode of perception. Listeners 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). In the present study, we refer to the amount of time that listeners must wait before making final decisions as memory load. The longer the interval between two stimuli or between stimulus and response, the more likely memory of the full range of phonetic details will have faded by the time the perceptual decision is made, i.e., high memory load. As a result, listeners will have to rely more on their longer-retained abstraction of phonological category information, i.e., more on a phonological mode of perception.

Consistent with that analysis, several studies have found that longer intervals between stimuli (Interstimulus Interval, ISI) lead to poorer discrimination of "difficult" non-native consonant (length) contrasts (Asano, 2018; Werker & Tees, 1984; Werker & Logan, 1985) that appear to fit PAM's definition of Single-Category assimilations, i.e., perceived as qualitatively equivalent phonetic variants within a single native category. Discrimination requires attention to phonetic differences between the stimuli in a trial and thus operates more on a phonetic mode of perception. Studies on discrimination of non-native consonants or consonant length contrasts have reported a switch from phonetic to phonological mode as a function of memory load. In one pioneering study on this issue, English participants completed a discrimination task with the "difficult" non-native Hindi voiceless unaspirated retroflex versus dental stop consonant contrast under two memory load conditions (Werker & Logan, 1985). They discriminated the contrast under low memory load (ISI = 500 ms), but not under high memory load (ISI = 1500 ms). Similarly, the performance of German native listeners decreased when discriminating non-native Japanese consonant length contrasts under high memory load, relative to low memory load (2500 vs 300 ms ISIs, respectively) (Asano, 2018).

However, the memory load effect appears to be different in tone discrimination. For example, when Mandarin listeners were asked to discriminate Cantonese tones or Cantonese listeners discriminated Mandarin tones, their very good performance was unchanged even under greatly lengthened ISI conditions (5 s) (Lee et al., 1996). Similarly, Yu and colleagues (2017) used a passive oddball paradigm with short versus long ISIs (600 vs 2600 ms) to test discrimination of Mandarin tones by native listeners and naïve English listeners who had no exposure to tone languages. Both behavioural and event-related brain potential data indicated no main effect of ISI for either the Mandarin or the English listeners, although the English group showed significantly lower discrimination than the Mandarin group in the long ISI condition. According to the cue-duration hypothesis (Fujisaki & Kawashima, 1970), the acoustic features responsible for distinguishing consonants, e.g., formant transitions or voice onset time, are generally short in duration and transient, and cannot be retained in detailed form in memory, whereas the phonetic details of longer-duration speech segments such as vowels can be better retained in memory. Thus, they argued that consonants are more likely to suffer from rapid memory decay than vowels. As the acoustic features of lexical tones extend over the entire sonorant portion of a syllable, i.e., the syllable's fundamental frequency contour, it appears that tones, like vowels, may also be retained in memory fairly well over time and if so, then mem-

ory load would also be expected to have little effect on non-native tone discrimination.

In contrast with the weak memory load effects on tone discrimination, given that categorisation tasks require processing the stimulus and accessing abstract phonological category information, we posit that a longer delay in responding to a target stimulus should provide the time needed for abstraction of categorical information, shifting listeners to a more phonological mode of perception. This has not been systematically examined in prior studies of non-native tone categorisation by tone-language listeners, nor in studies of non-native consonant or vowel categorisation. We hypothesise that in categorisation, the phonetic details of a non-native tone are initially stored in working memory. During this time, the phonological categories of the listeners' native tones that are most phonetically similar to the non-native tone are activated to be compared with the incoming non-native tone. Listeners will abstract higher level phonological information from the target stimulus and retain that abstract information when required to delay responding rather than relying on the phonetic details that are still available in working memory when pushed to respond more quickly. Thus, longer response intervals in categorisation will increase memory load and induce stronger native phonological influences. In addition, we assume that Categorised assimilation indicates stronger native phonological influence than UnCategorised assimilation. Thus, we propose that a non-native phone that is UnCategorised under low memory load, indicating weaker native phonological influence and more phonetic-level processing, could change to Categorised under high memory load, indicating strong native phonological influence due to the fading of non-native phonetic details in memory and more phonological abstraction.

Nonetheless, we propose that listeners do retain some residual sensitivity to within-category phonetic variations from their native categories in categorisation tasks, which will be reflected in their percent choice and goodness ratings (Chen et al., 2020). High residual sensitivity is reflected in low percent choice and category-goodness ratings whereas low residual sensitivity is reflected in high percent choice and ratings. When memory load is low and listeners rely on low level phonetic information, listeners are expected to show high residual sensitivity (lower percent choice and ratings) to within category phonetic variations when categorising non-native tones to their native lexical tones. When memory load is high and listeners rely more on abstract phonological information, listeners should, conversely, show low residual sensitivity (higher percent choice and ratings) instead.

1.2.2. Stimulus variability: Talker and phonetic context differences

Extending principles from ASP, when dealing with substantial linguistically irrelevant variability in speech, listeners adopt a mode of perception where they focus on abstract phonological information. On the other hand, listeners are more likely to use a phonetic mode of perception when variability of speech is low because phonetic information is constant and reliable in this case, and no abstraction across variability is needed. Two most common types of variability are talker and phonetic context variabilities. Two talkers produce the same segment with different acoustic–phonetic details due to differences in the shape and length of their vocal tracts and other indexical char-

acteristics that affect speech production (Nusbaum & Morin, 1992). Listeners need to somehow discount such linguistically irrelevant talker-specific variations to perceive constant, abstract phonological categories across different talkers. Talker variability affects speech perception: poorer performance in consonant and vowel recognition (Nusbaum & Morin, 1992) and identification (Mullennix, Pisoni, & Martin, 1989), and lexical tone identification (Wiener & Lee, 2020; Wong & Diehl, 2003). That is, there is some processing cost to abstracting across phonetic variation.

Similarly, varying phonetic contexts of a target phoneme also lead to phonetic variations in that phoneme and consequently affect its perception. Lexical tones are affected by vowel contexts. Other things being equal, high vowels, such as [i], have a higher intrinsic f₀ than low vowels, such as [a], in a number of languages (Ewan, 1975; Whalen & Levitt, 1995). Consistent with these intrinsic f₀ effects, the time-course of native tone language speakers' recognition (Shaw & Tyler, 2020) and categorical perception of tones (Zheng, 2014) is affected by vowel context. However, the effects of vowel context variability on categorisation and discrimination of non-native tones relative to a constant phonetic context have not been examined.

We argue that listeners generally process the stimulus variabilities via a similar mechanism as they categorise the indexical properties of unfamiliar talkers to the key phonological categories of their native speech community without removing irrelevant phonetic information (Best, 2015) and maintain phonological constancy. We hypothesise that the stimulus variability should affect categorisation and discrimination differently. In cross-language categorisation, participants are required to use their native phonological categories and to abstract more phonological properties from non-native tones to decide which category label to select and are prone to use a phonological mode of perception. Their categorisation judgments should be less susceptible to stimulus variability than their discrimination performance.

On the other hand, in discrimination, listeners are required to attend to phonetic differences between stimuli and thus engage in a more phonetic mode of perception by default. However, high stimulus variability in a discrimination task should nonetheless render specific phonetic details less reliable to accomplishing the judgment and thus lead listeners to rely more on abstract phonological information, i.e., to use a more phonological mode of perception. There is a processing cost to this abstraction across variations, which should be reflected in their response patterns, e.g., reduced accuracy in discrimination. There should also be variations in discrimination accuracy among differently assimilated contrasts. In high variability conditions, accuracy should be reduced especially for non-native contrasts within the same native phonological category, e.g., Single-Category contrasts and to a lesser extent Category-Goodness contrasts, and for those with high degree of phonological and/or phonetic overlap.

1.3. Lexical tones in Thai, Mandarin and Vietnamese

We relied on existing Chao transcriptions (Chao, 1930) to provide an initial approximation of the phonetic characteristics of the tones in our selected languages. Chao transcriptions are

based on perceived pitch within speakers' vocal range, in which 5 represents the highest and 1 the lowest pitch in that range. Chao transcriptions are somewhat more idealised than actual acoustic f0 patterns, but not as abstract as phonological features. Thus, the Chao numbers sometimes differ a bit from the specific, concrete f0 acoustic properties, as show in Fig. 1, as well as from the more abstract phonological features of tones, i.e., level, rising, falling, rising-falling, falling-rising and heights, i.e., high, mid, low. Thai (T) contrasts five lexical tones: three phonologically level tones, high-level, mid-level and low-level; and two contour tones, rising and falling (Gandour, 1978). Two of the three phonologically level tones in Thai are not *phonetically* flat, as can be seen in both the acoustics (left panel of Fig. 1) and in the Chao transcriptions: high-level is transcribed as T45 and shows a moderate f0 rise, low-level as T21 and shows an f0 fall, while only mid-level T33 is phonetically flat according to the Chao transcription and nearly so in its acoustic pattern (Reid et al., 2015). When produced in isolation, the phonologically rising tone is transcribed as T315 and its f0 shows a falling portion at its onset, while the phonologically falling tone is transcribed as T241 and shows a short rise in f0 at its onset.

Mandarin (M) has four tones (Chao, 1968), of which one is phonologically level and three are phonologically contoured in their citation forms: a high-level tone M55 (called Tone 1 in Mandarin) which shows a slight f0 rise at the onset; a rising tone M35 (Tone 2) which shows a nearly flat f0 onset before it rises; a falling-rising tone M214 (Tone 3); and a falling tone M51 (Tone 4) which shows a brief f0 rise at onset before it falls (see Fig. 1, middle panel).

Vietnamese¹ has five tones in its phonological system (Brunelle, 2009): two phonologically level tones, high-level V44, also called *ngang* in Vietnamese, and low-level V22 (*huyền*) which both show modest f0 falls; and three phonologically contour tones: rising V35 (*s c*), falling V21 (*nặng*), which appears to have a modest final f0 rise, and falling-rising V214 (Nhàn, 1984). V214 results from a merger of the two falling-rising tones in the standard/Northern dialect, NV214 (*hỏi*), and NV415 (*ngã*) (Fig. 1, right panel).

In this paper, we first report an experiment that examined how two factors, i.e., memory load and stimulus variability, biased listeners of two tone languages toward a more phonological versus phonetic mode in their categorisation of non-native tones to their native tone categories. Native phonological and phonetic influences on tone perception were expected to be modulated by the two factors in affecting listeners' categorisation patterns, i.e., assimilation types and category-goodness ratings. In the second experiment, we also manipulated memory load and stimulus variability in discrimination and tested PAM predictions about discrimination of the same five non-native tone contrasts which were assimilated differently for each listener group using their categorisation patterns. Their discrimination performances were expected to reflect both native language phonological and phonetic influences as indicated by specific categorisation patterns and should also be modulated by memory load and stimulus variability.

2. Experiment 1: Mandarin listeners' perception of Thai tones

Experiment 1 was designed to examine perceptual assimilation and discrimination of Thai tones by Mandarin listeners. A within-subjects design with blocked memory load conditions, i.e., high versus low, would have granted greater statistical power in data analysis. However, we decided against within-subjects as participants are unlikely to switch between two modes of perception within an experiment, and if required to do so they tend to rely on strategies during the second part of the study that they had developed in the prior test blocks (Werker & Logan, 1985). For this reason, a between-subjects design was employed for the memory load conditions.

Five Thai contrasts were used in the discrimination task, selected based on findings from a previous study with these stimulus and listener languages (Chen et al., 2020) to present a rich range of assimilation patterns and predicted differences between the two listener languages. In that study, Mandarin listeners Categorised T45 to their native M35, T33 to M55, T21 to M214 and T315 to M35. T241 was an UnCategorised assimilation split between M55 and M51. For the present study, we selected five Thai contrasts T33-T21, T33-T45, T33-T241, T21-T241, and T45-T315 based on the prior study in which Mandarin listeners displayed Two-Category assimilations for T33-T21 and T33-T45; UnCategorised-Categorised assimilation for T33-T241 and T21-T241; and Single-Category assimilation for T45-T315. However, in that study the categorisation response intervals were not manipulated, and participants were instructed simply to respond as quickly as possible. As the response intervals were manipulated in the categorisation task of the present study, we expected that the assimilation results could differ somewhat from the untimed responses of the prior study. We therefore based our discrimination predictions for the current study on the assimilation results we obtained here using manipulations of the response interval.

2.1. Method

2.1.1. Participants

Mandarin speakers ($n = 32$) were divided equally into two groups for the two memory load conditions (see *Procedure*) (Low: $M_{\text{age}} = 26.6$ years, $SD = 7$ years, 10 females; High: $M_{\text{age}} = 26.5$ years, $SD = 6.4$ years, 10 females). The participants were living in Sydney at testing time, but all were born and raised in various regions in China (i.e., Tianjin, Anhui, Sichuan, Shandong, Henan, Hunan, Jilin, Jiangsu, Jiangxi, Shanxi, Xinjiang, Shanghai, Inner Mongolia)² and educated in Mandarin from early childhood through high school, and used Mandarin on a daily basis. None had more than two years of formal musical training, which is known to influence tone perception (Gottfried, Staby, & Ziemer, 2004). All reported normal hearing, and none had experience with Thai. The experiments were approved by the Western Sydney University Human Research Ethics Committee and all participants gave informed consent prior to testing.

² In future research, more homogeneous groups of participants would be desirable, e.g., recruiting only Beijing Mandarin speakers. For practical reasons, however, strict control of dialect background of Mandarin participants in Sydney was not possible. With this said, the possible effect of dialect background differences adds potential variation to the assimilation patterns. In the present study, we used participants' assimilation patterns to directly predict and account for variations in their discrimination accuracy. In this way, dialect effects should be consistent across assimilation and discrimination tasks.

¹ Our Vietnamese listeners all spoke the Southern Vietnamese dialect.

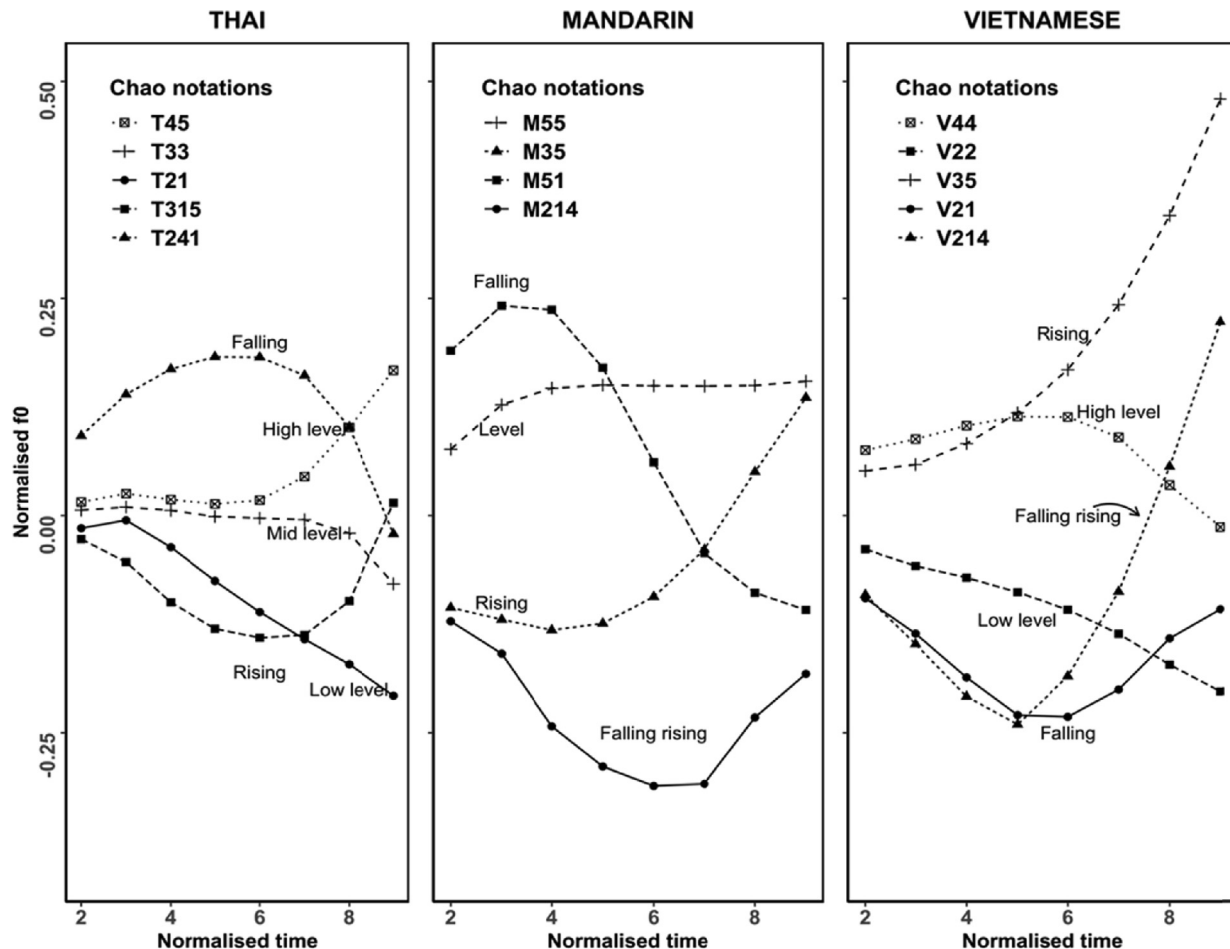


Fig. 1. Time- and Lobanov-normalised (Lobanov, 1971) f0 contours of Thai, Mandarin and Southern Vietnamese tones (In Chen et al (2020), Southern Vietnamese speakers produced both V214 (hỏi) and V415 (ngã) with no significant acoustic differences, consistent with the reports of merger, so here they were averaged and labelled as a single phonologically falling rising tone SV214.). The legends in each panel show the Chao notations for the tones of the respective languages.

2.1.2. Stimulus materials

Two syllables (/ma/, /mi/) were chosen for target stimuli because they are phonotactically legal and form meaningful morphemes for each native tone in Thai and Mandarin. Thus, Thai-naïve Mandarin listeners were able to categorise the Thai tone stimuli into their native phonological system as morphemes. The Thai syllables were recorded for a separate study (Burnham, Kuratate, McBride-Chang, & Mattock, 2009), and were used here with permission from those authors. The target Thai syllables were each read several times by two female native Thai speakers who had no experience with other tone languages. Two tokens of each target item per speaker were selected and judged to be correct and most natural sounding to a third native Thai speaker.

In all the following experiments, we systematically manipulated two sources of stimulus variability: number of talkers (one vs two) and/or vowel contexts (one vs two). Tone realisation is affected both by f0/contour differences between talkers and by intrinsic f0 differences between vowels, which is higher for the high vowel [i] in /mi:/ than the low vowel [a] in /ma:/ (see Fig. 2).

2.1.3. Procedure

Participants were tested individually in a quiet testing space. Stimuli were presented on a Dell Latitude 7280 laptop running E-Prime Professional 2, via Sennheiser HD 280 Pro

headphones at 72 dB SPL. In all experiments, discrimination tests were run before categorisation tests to minimise the influence of categorisation on discrimination decisions. Timing intervals were held constant across the two tasks within each memory load condition. The low memory load condition used 500 ms as the interstimulus interval (ISI) in the discrimination task and as the Response Interval (RI) in the categorisation task, to ensure the availability of phonetic details in working memory (Asano, 2018). The high memory load condition used 2000 ms ISIs and RIs, respectively, to optimise adequate decay of phonetic details (i.e., even larger than the 1500 ms intervals in Werker & Tees, 1984; Werker & Logan, 1985).

2.1.3.1. Discrimination. An AX discrimination task (“same-different”) was used because it allows for a single ISI interval between the stimuli to be judged, thus offering a clearer interpretation of memory load effects than tasks such as AXB or 4AFC which respectively require two or three ISI intervals on each trial. The AX trial format has been used in many previous discrimination studies involving ISI manipulations (Asano, 2018; Werker & Tees, 1984; Werker & Logan, 1985). On each trial, participants heard a stimulus A followed by a silence of either 500 or 2000 ms. After the ISI, the second stimulus X was presented and “Same or different?” appeared on the screen to signal to the participant that they should respond. They were

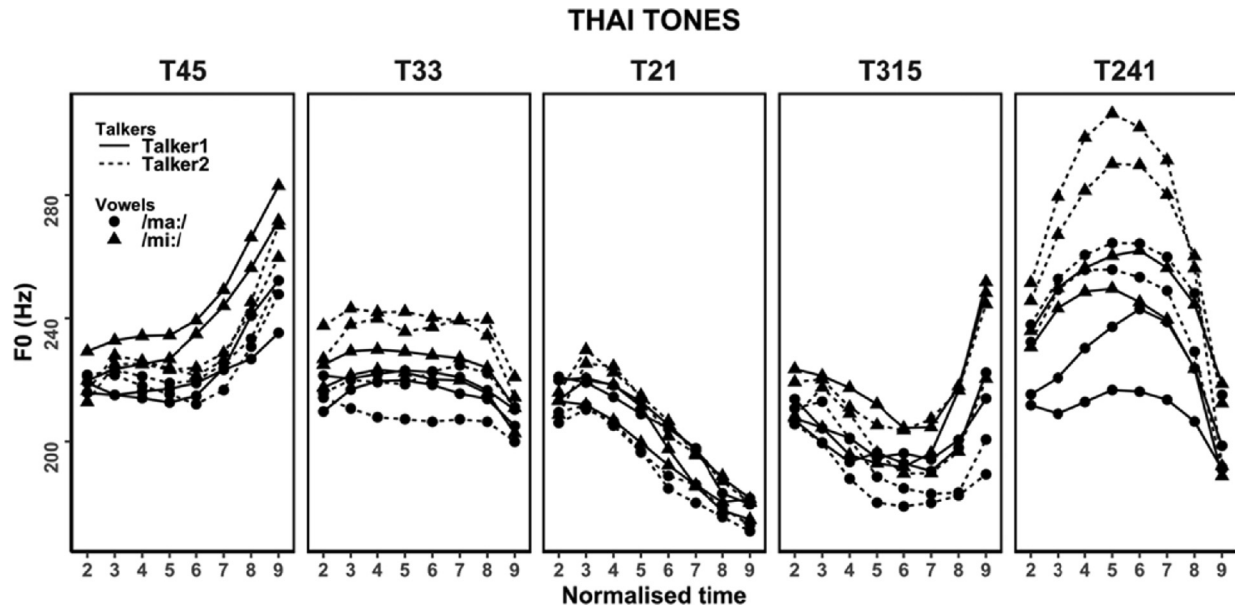


Fig. 2. Time-normalised f0 contours of the five Thai tones produced by the two native female informants with /ma:/ and /mi:/ (each with two tokens per target per speaker that were used in the perceptual experiment).

told to only consider the tone. Half of the trials were “same tone” trials and the other half were “different tone” trials; these were presented in random order.

Participants were given 3 seconds to answer before timeout on each trial, and the inter-trial interval following their response (or timeout) was 1 second. In each memory load condition, both talker and vowel variability conditions were manipulated across eight randomised blocks (Table 1), i.e., Constant vs Variable: one versus two talker/vowel contexts, respectively. Within each testing block, each of the resulting stimulus pairings was repeated two times each, in random order, for the four AX trial types, i.e., for stimulus A and B: AA, AB, BA, BB. Note that in the same talker same vowel condition, we used two tokens of the same syllable produced by the same speaker, which display phonetic differences (Werker & Logan, 1985). Including token variability of the same speaker should tap into discrimination mostly at the phonetic level although it will involve some phonological abstraction due to the token variation but at a much lesser level than multiple talkers and/or different vowel contexts. Before the discrimination test, participants completed 16 practice trials with same syllables produced by a third Thai speaker which were not used in the main experiment.

2.1.3.2. Categorisation. On each trial, a stimulus token was presented, and listeners made a forced-choice categorisation judgment to their native tones via a keypress to one of four Pinyin options. The stickers were placed on the keys, i.e., “f”, “g”, “h”, “j”, in the same line on the keyboard. They were instructed to press the key only after they saw “which tone?” onscreen; they had a 3-second time limit after the display. The interval between stimulus presentation and the signal to answer (“which tone?”), i.e., response interval or RI, was 500 ms under low memory load and 2000 ms under high memory load, consistent with our manipulation of ISI in the discrimination task. We note that the signal to respond, “Which tone?”, in the present experiment did not actively prevent listeners from deciding the tone category response before the “which tone?” signal, but it did prevent them from *responding* until they received the signal to go. Immediately after their response, they heard the tone again and rated its goodness of fit to their chosen native category on a 7-point scale via a keypress: 1 = poor, 7 = perfect. Talker and vowel variability were manipulated across nine blocks (see Table 2). Each tone was tested under each stimulus variability condition with two tokens from each speaker except for the variable-talker-variable-vowel block, in which ten /ma:/ trials (5 tones × 2 talkers) and ten /

Table 1

Talker and vowel variability conditions and AX pair block types within each condition for the discrimination task (4 conditions, 40 trials per block, 8 blocks).

Conditions	Blocks
Constant Talker + Vowel	Talker 1 /mi:/ + /mi:/ and /ma:/ + /ma:/ Talker 2 /mi:/ + /mi:/ and /ma:/ + /ma:/
Constant Talker/ Variable Vowels	Talker 1 with /ma:/ + /mi:/ and /mi:/ + /ma:/ Talker 2 with /ma:/ + /mi:/ and /mi:/ + /ma:/
Variable Talkers/ Constant Vowel	Talker 1 + Talker 2 and Talker 2 + Talker 1 /ma:/ Talker 1 + Talker 2 and Talker 2 + Talker 1 /mi:/
Variable Talkers + Vowels	Talker 1 /ma:/ + Talker 2 /mi:/ and Talker 1 /mi:/ + Talker 2 /ma:/ Talker 2 /ma:/ + Talker 1 /mi:/ and Talker 2 /mi:/ + Talker 1 /ma:/

Table 2

Talker and vowel variability conditions and block types within each condition for the categorisation task. Each block in Constant Talker + Vowel had 10 trials; blocks in the remaining three conditions had 20 trials each. Trials were randomised within blocks.

Conditions	Blocks
Constant Talker + Vowel	Talker 1 /mi:/ Talker 1 /ma:/ Talker 2 /mi:/ Talker 2 /ma:/
Constant Talker/ Variable Vowels	Talker 1 with /ma:/ + /mi:/ Talker 2 with /ma:/ + /mi:/
Variable Talkers/ Constant Vowel	Talker 1 + Talker 2 with /ma:/ Talker 1 + Talker 2 with /mi:/
Variable Talkers + Vowels	Talker 1 + Talker 2 with /ma:/ + /mi:/

mi:/ trials (5 tones $\times$ 2 talkers) from one token were included, to avoid participant fatigue and keep the duration of the experiment reasonable. The differences between two tokens of the same talker and vowel should be relatively small compared to talker and vowel variations. However, this condition had a smaller number of data points ($n = 20$) relative to other conditions ($n = 40$), which reduced some of its statistical power but for the categorisation task only. In addition, this would not confound the memory load effect on categorisations, as participants in both memory groups did the same task. Before the categorisation test, participants completed 10 practice trials with same syllables produced by a third Thai talker which were also used for the AX practice trials but not used in the main experiment.

2.2. Results

The categorisation results are presented first because they were used to determine the assimilation types for each of the five Thai tone contrasts and form the basis for PAM predictions about discrimination.

2.2.1. Categorisation

Table 3 shows the mean percent choice of categorisations to Mandarin tones and goodness of fit ratings for each Thai tone. Missing data, i.e., responses before the signal “Which tone?” or responses after timeout accounted for 0.53 % of all the data and were deleted before analysis. Although in some studies on consonant and vowel perception, Categorised assimilation has been defined as the selection of one native category (Tyler, Best, Faber, & Levitt, 2014) for a given non-native phone above a fixed threshold (e.g., 70 %), we followed two statistically-based criteria used in So and Best (2014), which are more sensitive to variations among different native tone systems than a fixed threshold would be, as they accommodate differences in the number of tone categories in listeners’ languages. First, a given native tone must be selected significantly above chance level as determined by a t -test: 25 % for Mandarin listeners with four native tone choices. Second, that same native tone category must be chosen significantly more often than any other response categories.

Nearly all choices were significantly above chance level (see Table 3), except for choices of M214 for T21 under low memory load and choices of M55 for T21 under high memory load (see Appendix A, Table A.1 for the full set of t -test results). In order to determine whether the above-chance response categories were chosen significantly more often than other response categories, we fitted the data using a Linear Mixed Effect Regression (LMER) model with percent choice as the dependent variable, native categories as a fixed factor, and subject as the random intercept. To calculate the p -values for the fixed effects, we used the Kenward-Roger approximation to the degrees of freedom, as recommended by Halekoh and Hojsgaard (2014), and the *Anova* function from the *car* package (Fox & Weisberg, 2019) in *R* (R Core Team, 2018), with test specified as “F”. We then ran multiple comparisons between native categories with the R-package *lsmeans* (Lenth, 2016). Following the above procedure, four Thai tones were deemed to have been Categorised to Mandarin tone categories by our Mandarin listeners: T21 was Categorised as

M51; T33 as M55; T45 as M35; and T241 as M51 (for statistical results see Appendix A, Table A.2 and Table A.3). T315 assimilation was UnCategorised and split between M35 and M214.

In order to test the effects of memory load and stimulus variability, we fitted the data with the *multinom* function³ in the R package *nnet* (Venables & Ripley, 2002). A full multinomial regression model of the categorisation data was conducted with Mandarin tone choices as a dependent measure, and memory load (Low and High), talker and vowel variability (Constant vs Variable) and Thai tones (T21, T33, T45, T241, T315) as fixed factors. Four additional models were built each with one fixed factor subtracted and these models were compared to the full model to determine the effects of cognitive load. Both memory load, $\chi^2(60) = 146.14$, $p < 0.001$, and Thai tone, $\chi^2(96) = 6342.04$, $p < 0.001$, showed significant effects in Likelihood ratio tests, suggesting that participants responded differently across the Thai tones and memory load conditions. However, the talker and vowel variability effects were non-significant.

2.2.2. Predictions for discrimination

The original PAM principles (Best, 1995) stated that discrimination of a Two-Category contrast should be excellent and that of an UnCategorised-Categorised contrast should be very good because both involve phones that sit on opposite sides of a native phonological category boundary. On the other hand, discrimination of a Category-Goodness contrast rests solely on within-category phonetic sensitivity without assistance from a native phonological difference and its discrimination may range from moderate to good depending on differences in perceived phonetic goodness of fit of each non-native phone to the native category.

More recently, it has been recognised that the degree of overlap in native category assimilations between contrasting non-native phones must also be considered (Best et al., 2019; Faris et al 2016, 2018; Tyler et al., 2014). When two non-native phones are perceived as phonologically similar to the same set of native categories, i.e., complete overlap, discrimination will be poor due to the strong interference arising from the identical chosen native phonological categories. In determining phonological overlap types, only above-chance native category choices were considered. When a non-native contrast is assimilated to sets containing *some but not all* of the same native categories, i.e., partial overlap, discrimination will improve but still be poorer than when the members of a non-native contrast are assimilated to *completely different* sets of native categories, i.e., non-overlap, in which native phonological distinctions facilitate discrimination. To illustrate, discrimination of an UnCategorised-Categorised/Partial-overlap contrast should be comparable to Two-Category/Partial-overlap but poorer than that of Two-Category/Non-overlap or Category-Goodness/Non-overlap.

³ This function does not cover mixed effects for subjects as a random factor. To incorporate subjects as a random factor we also ran a Bayesian Markov Chain Monte Carlo (MCMC) analysis using the *MCMCglmm* package (Hadfield, 2010). The p values for significant effects are pMCMC values. That model showed similar results as the multinomial model, i.e., a significant effect for ISI and tone types but not for talker and vowel variability. However, the result was not stable even with a large number of iterations due to the random sampling aspect of the method. To be consistent with later models, we report the more stable multinomial model results here.

Table 3

Assimilation of Thai tones into Mandarin tone categories under low versus high memory loads. Categories in bold are choices that were significantly above chance: 25 % for Mandarin; *** = Categorical tone. Assimilations: C = Categorical, U = UnCategorised. Rating: 1 = poor, 7 = perfect; mean ratings are displayed. "-" = no responses.

Memory Load		Thai stimulus									
		T45		T33		T21		T315		T241	
Low	Response	%	rating	%	rating	%	rating	%	rating	%	rating
	M55	0.4	4.0	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
	M214	10.3	4.3	0.7	3.3	25.8	3.7	51.2	5.4	0.4	3.5
	M51	0.4	5.5	19.2	4.9	52.7*	4.6	0.2	7.0	75.1*	5.6
Assimilation		C		C		C		U _{clustered}		C	
High	M55	-	-	84.7*	5.2	26.4	3.2	0.2	7.0	28.1	5.0
	M35	85*	5.3	0.2	5.0	1.1	3.3	44.2	5.1	0.2	2.0
	M214	14.7	4.8	1.7	5.9	6.3	3.8	55.6	5.5	0.5	6.0
	M51	0.2	2.0	13.4	4.4	66.2*	4.0	-	-	71.2*	5.1
Assimilation		C		C		C		U _{clustered}		C	

First, we identified phonological assimilation types and overlap types for our selected pairs of Thai tones, based on the current perceptual categorisation results. The manipulation of memory load in the categorisation task affected some, but not most, assimilation and overlap types. T33-T45 and T33-T21 were both Two-Category contrast assimilations with no overlap, i.e., Two-Category/Non-overlap, under both memory loads. T33-T241 was a Two-Category assimilation but with partial overlap, i.e., Two-Category/Partial-overlap, under high memory load. Under low memory load, T33-T241 remained a Two-Category/Non-overlap assimilation. T315-T45 was an UnCategorised-Categorised/Partial-overlap assimilation with one overlap in Mandarin categories chosen under both memory loads. Given that both T21 and T241 were assimilated to M51, we ran *t*-tests for each memory load condition to determine whether the T21-T241 contrast assimilation was a Single-Category or Category-Goodness type. The rating differences were significant, i.e., low memory load: $t(25) = -2.47$, $p = 0.02$; high memory load, $t(28) = -2.67$, $p = 0.01$, indicating Category-Goodness assimilation under both memory loads. Thus, we did not find any Single-Category contrast assimilations in this experiment.

To indicate the differences between two memory loads, we provide two sets of PAM assimilation type/overlap type predictions accordingly, one per memory load condition. Under low memory load, discrimination accuracy predictions are, from best to worst: Two-Category/Non-overlap {T33-T45 = T33-T21} > Category-Goodness/Non-overlap {T241-T21} > UnCategorised-Categorised/Partial-overlap {T315-T45}. On the other hand, under high memory load, T33-T241 became Two-Category/Partial-overlap. Consequently, predicted discrimination accuracy from best to poor worst was: Two-Category/Non-overlap {T33-T45 = T33-T21} > Category-Goodness/Non-overlap {T241-T21} > Two-Category/Partial-overlap {T33-T241} = UnCategorised-Categorised/Partial-overlap {T315-T45}.

In addition, for non-native contrasts of the same assimilation and phonological overlap type, we reasoned that listeners' phonetic sensitivity to contrasts can vary as a function of perceived phonetic similarities of the contrasting tones, as indicated by percent choice and goodness ratings. For example, the Two-Category/Non-overlap contrasts, such as T33-T45

and T33-T21, might differ in their perceived phonetic overlap, although their phonological overlap types were the same, according to only above-chance native category choices.

One approach for considering residual phonetic sensitivity is by computing *overlap scores* (as in Flege & MacKay, 2004; Levy, 2009) as the summed percentages of all overlapping native category choices, including below chance choices, for a non-native contrast. For example, using the group-level data from Table 3 for T33-T21, percent of each overlapping native category choice should be 0.20 for M55, 0.02 for M35, 0.007 for M214, 0.19 for M51 and the overlap score should be the sum of all the percentages, 0.42. Note that, in practice, we calculated overlap scores for each participant and based our analysis on the group mean of overlap scores.

However, Goodness ratings were not considered in those previously-developed overlap score formulae. Given that both goodness ratings and percent choice indicate residual phonetic sensitivity to variations of the non-native stimulus from the chosen native category (Chen et al., 2020), we also calculated a *fit-index* (Wu et al., 2014) by multiplying percent choice by goodness ratings (see Appendix A, Table A.4). To illustrate, we again use the group-level data from Table 3 for T33-T21. First, the fit-index of each native category was calculated for each Thai tone, e.g., $4.10 (=0.773 \times 5.3)$ for T33 assimilated as M55 and $0.66 (=0.199 \times 3.3)$ for T21 assimilated as M55. Then the absolute value of difference between the two fit-index scores, i.e., $4.10 - 0.66$, was calculated and the fit-index difference score was generated, 3.44. This was done for each native response category and summed as the final fit-index difference score for the contrast. Note that for our actual analyses we normalised the goodness rating for each participant, i.e., $(\text{rating} - \text{rating}_{\min}) / (\text{rating}_{\max} - \text{rating}_{\min})$, so as to reduce individual variation and to make the ratings comparable. This yielded smaller fit-index difference scores than in our group-based illustration here. A high overlap score between contrasting non-native phones indicates *low* listener sensitivity to phonetic differences between them. Conversely, a high fit index difference between contrasting non-native stimuli, or the *fit index difference score*, represents *high* residual phonetic sensitivity to differences between them.

In the present study, we used the confidence intervals around the overlap scores and fit-index difference scores of

contrasts with the same assimilation and/or phonological overlap type to indicate their phonetic overlap degrees and predict differences in their discrimination. For example, T33-T45 and T33-T21 were both assimilated to Mandarin tones as Two-Category/Non-overlap. The overlap scores were lower and the fit indices higher for T33-T45 ($M_{\text{Overlap_score}} = 0.039/0.009$; $M_{\text{Fit_index}} = 1.75/1.76$) than for T33-T21 ($M_{\text{Overlap_score}} = 0.36/0.41$; $M_{\text{Fit_index}} = 1.07/0.99$) under both memory loads. By the reasoning we provided above, rather than discrimination performance being equivalent for the two contrasts, T33-T45 = T33-T21, as predicted earlier, differences in residual phonetic sensitivity for these two non-native tone contrasts should result in differential discrimination: T33-T45 > T33-T21. This is the only change in the predictions. All other predictions listed earlier remained unchanged due to lack of differences in our measures of residual phonetic sensitivity.

2.2.3. Discrimination

In order to minimise response bias effects on the AX data, we calculated d' (Macmillan & Creelman, 2005) for discrimination of each tone pair in each cognitive condition, with adjustments made for probabilities of 0 to avoid infinite values, by adding 0.5 to all raw data cells regardless of whether zeroes are present (Hautus, 1995; Miller, 1996). Hit is defined as the number of correct “different” responses on AB or BA trials. False positive is defined as the number of incorrect “different” responses on AA or BB trials. The d' scores were calculated using a differencing rule for AX tasks (as in Macmillan & Creelman, 2005) for each participant and separately for each tone pair and for each block. 0.55 % of all the discrimination data were time-outs and were removed as missing data. We fitted the data using an LMER model (via *lme4*, Bates, Mächler, Bolker, & Walker, 2015) with d' as the dependent variable, and memory load (Low and High), talker and vowel variability (Constant vs Variable) and tone contrasts (T241-T21, T33-T21, T315-T45, T33-T241, T33-T45) as fixed factors, and participant as a random factor including random intercepts and slopes for within-subject fixed factors, i.e., talker and vowel variability conditions to make the model maximally generalizable across the data (as suggested by Barr, Levy, Scheepers, & Tily, 2013).⁴ There are several different methods of estimating the significance of fixed factors and calculating relevant p -values for the fixed effects in LMER models (Luke, 2017). We used the Kenward-Roger degrees of freedom approximation (Halekoh & Hojsgaard, 2014), and the *Anova* function from the *car* package in R, with test specified as “F” (see Table 4, for the coefficients of the fixed effects and interactions see Table A.5 in the Appendix A). Here we report main effects for multilevel factors or interactions, 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 talker variability, vowel variability, and Thai tone contrasts, and interaction of talker variability $\times$ vowel variability. However, the main effect of memory load and all interactions involving memory load were non-significant. The scores for the constant-talker blocks ($M = 3.26$, 95 % CIs [3.11, 3.41]) were significantly higher, suggesting better discrimination, than those of the variable-talker blocks ($M = 2.79$, 95 % CIs [2.64, 2.94]). Similarly, the scores for constant-vowel blocks ($M = 3.11$, 95 % CIs [3.16, 3.46]) were significantly higher than those of variable-vowel blocks ($M = 2.74$, 95 % CIs [2.59, 2.89]).

To further examine the Thai tone contrast main effect and test PAM predictions, we ran pairwise multiple comparisons with Tukey adjustments for each tone pair contrast using the *lsmeans* package (see Appendix A Table A.6 for full statistical details). This method is widely used and accepted without having to choose an arbitrary reference level (Tang, Yuen, Xu Rattanasone, Gao, & Demuth, 2019) and also is a good fit for evaluating our theoretical predictions and experimental design.

All contrast comparisons were significant, except for the T315-T45 versus T33-T241 comparison. These results reflect the following pattern: T33-T45 ($M = 4.33$, 95 % CIs [4.15, 4.52]) > T33-T21 ($M = 3.42$, 95 % CIs [3.21, 3.63]) > T241-T21 ($M = 3.05$, 95 % CIs [2.85, 3.26]) > T33-T241 ($M = 2.06$, 95 % CIs [1.87, 2.25]) = T315-T45 ($M = 2.26$, 95 % CIs [2.06, 2.45]).

As noted earlier, residual phonetic sensitivity differences between the two Two-Category/Non-overlap assimilation contrasts predicted differential discrimination consistent with the relevant observation in our pairwise comparisons: T33-T45 > T33-T21.

In addition to main effects, multiple comparisons were conducted with Tukey adjustments to break down the talker variability $\times$ vowel variability interaction (see Fig. 3). In the constant-talker condition, the constant-vowel blocks ($M = 3.71$, 95 % CIs [3.51, 3.90]) were discriminated better than the variable-vowel blocks ($M = 2.81$, 95 % CIs [2.60, 3.03]), $\beta = 0.90$, $SE = 0.11$, $t(108) = 8.55$, $p < 0.001$. Similarly, in the constant-vowel condition, the constant-talker blocks ($M = 3.71$) were discriminated better than the variable-talker blocks ($M = 2.91$, 95 % CIs [2.70, 3.12]), $\beta = 0.80$, $SE = 0.10$, $t(113) = 7.70$, $p < 0.001$. In addition, the constant-talker + vowel block had significantly higher scores ($M = 3.71$) than the variable-talker + vowel block ($M = 2.67$, 95 % CIs [2.46, 2.88]), $\beta = 1.04$, $SE = 0.11$, $t(60) = 9.90$, $p < 0.001$.

2.3. Discussion

In this experiment, our PAM-based predictions with consideration of phonological overlap types and phonetic overlap scores were upheld. The Two-Category/Non-overlap contrasts (T33-T45 and T33-T21) were better discriminated than Category-Goodness/Non-overlap contrast (T241-T21). Within the same assimilation-overlap category, both overlap scores and fit index difference scores successfully predicted T33-T45 to be better discriminated than T33-T21.

Native phonological category overlap clearly affected non-native tone discrimination. Discrimination of T33-T241, a

⁴ Regarding model selection, we did not remove random intercepts or slopes although doing so improved the model as indicated by a smaller AIC value, but no significant improvement as indicated by Likelihood ratio tests in our case. For fixed factors, we followed the advice from two anonymous reviewers, and rebuilt the models with the non-significant interactions removed. The original significant results remained significant. In other words, removing those non-significant interactions does not affect the significant results nor our interpretations from these results. We chose to report the full model so that readers have information on all factors and interactions. It also fits the experiment design better, as these factors were all carefully selected and need to be considered.

Table 4

Main effects and interactions for Thai tone contrast discrimination by Mandarin listeners. “*” indicates significant.

Cognitive factors	<i>F</i>	<i>df</i>		<i>p</i>
Memory load	0.08	1	30	0.78
Talker variability	41.08	1	30	< 0.01*
Vowel variability	57.40	1	30	< 0.01*
Tone contrasts	127.23	4	510	< 0.01*
Memory load × Talker variability	0.04	1	30	0.84
Memory load × Vowel variability	0.02	1	30	0.88
Talker variability × Vowel variability	20.06	1	510	< 0.01*
Memory load × Tone contrasts	0.31	4	510	0.87
Talker variability × Tone contrasts	1.55	4	510	0.19
Vowel variability × Tone contrasts	1.03	4	510	0.39
Memory load × Talker variability × Vowel variability	0.41	1	510	0.52
Memory load × Talker variability × Tone contrasts	0.05	4	510	1.00
Memory load × Vowel variability × Tone contrasts	0.91	4	510	0.46
Talker variability × Vowel variability × Tone contrasts	0.33	4	510	0.86
Memory load × Talker variability × Vowel variability × Tone contrasts	0.12	4	510	0.98

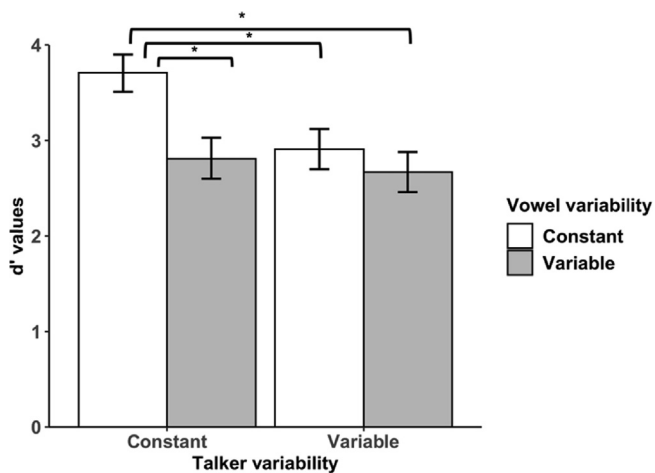


Fig. 3. Interaction between talker and vowel variability in discrimination by Mandarin listeners. d' plotted along the y axis was calculated using a differencing rule for AX tasks (Macmillan & Creelman, 2005). Error bars indicate 95% confidence intervals around the mean.

Two-Category/Partial-overlap contrast under high memory load, was worse than those of both Two-Category/Non-overlap contrasts, T33-T45 and T33-T21, and was as poor as that of UnCategorised-Categorised/Partial-overlap contrast T315-T45, as predicted. In addition, T315-T45 was also discriminated with lower accuracy than T241-T21, Category-Goodness/Non-overlap, as predicted due to phonological overlap.

Categorisation and discrimination performance were affected differently by our memory and variability manipulations. Memory load affected percent choice in the categorisation task. As we have previously suggested (Chen et al., 2020), Categorised assimilations reflect a clear *phonological* influence from the native language tone system, whereas UnCategorised assimilations reflect weaker native phonological influence. In addition, high percent choice and/or goodness ratings for a Categorised native response reflect low residual phonetic sensitivity to the difference of the non-native category from the native category. The percent choice of T33 increased for the Categorised native response M55, the only Mandarin level tone, under high memory load, suggesting stronger native phonological constraints. For the Categorised high level Thai tone T45 and falling tone T241, the differences were small

between the two memory load conditions. The UnCategorised Thai tone T315 also showed small differences between the two memory load conditions as expected, in this case because UnCategorised assimilations reflect weaker native phonological influence. On the other hand, there was a sizeable difference in T21 between two memory load conditions (around 14 %). T21 was Categorised more consistently under high memory load.

Talker and vowel variability did not affect assimilation. We reason that as categorisation explicitly requires participants to use their native phonological categories, they were likely to use a more phonological mode of perception. Thus, according to ASP (Strange, 2011), they would have been attending to abstract phonological information when making judgements and were less susceptible to linguistically irrelevant phonetic variability, i.e., talker and/or vowel variability. We further argue that the phonetic variabilities are accounted for by a similar mechanism as in cross-language categorisation because perceivers can display phonological constancy across the indexical properties of unfamiliar talkers for the phonological categories of their native speech community (Best, 2015). In other words, for categorisation, Mandarin listeners can rely on mechanisms used in their native language to maintain phonological constancy, i.e., the ability to keep word and lexical tones' phonological identity intact, and accommodate to lexically irrelevant variability.

On the other hand, discrimination was unaffected by memory load, differing from prior work with consonants (Werker & Logan, 1985), but consistent with previous findings on tone perception (Lee et al., 1996). This finding is in line with the cue-duration hypothesis that the longer the duration of an acoustic property, the less likely it is to decay in short-term memory (Fujisaki & Kawashima, 1970). The acoustic properties of lexical tones, i.e., fundamental frequency and contour, extend over the entire duration of sonorant syllables, i.e., [ma:] and [mi:], should be less likely to decay, and therefore should be more stable in memory than the more transient acoustic properties of consonants, e.g., formant transitions or voice onset time.

Both talker and vowel context variability affected discrimination of non-native tones, unlike their lack of significant impact on assimilation. In the assimilation task, more attention was dedicated to grouping perceptually similar objects into the same category at a more abstract phonological level, whereas

in the discrimination task listeners were required to attend to phonetic differences and thus they should be biased toward a phonetic mode by default. Nonetheless, high stimulus variability should bias listeners to a more phonological mode even in discrimination, requiring them to perceive abstract pitch contours and relative heights and reducing their attention to concrete phonetic f_0 differences, some of which were important to discriminating between tones. Consequently, their discrimination accuracy was reduced.

3. Experiment 2: Vietnamese speakers' perception of Thai tones

Experiment 2 tested whether the effects observed in Experiment 1 extend to listeners of another tone language that has a different tone system than both Mandarin and Thai. Vietnamese was selected as it satisfies these requirements. This enabled us to test whether the same five Thai tone pairs, i.e., T241-T21, T33-T21, T315-T45, T33-T241 and T33-T45, would form different contrast assimilation patterns in Vietnamese than in Mandarin, and in turn lead to language group differences in discrimination performance in accordance with PAM-based predictions. For Vietnamese listeners, according to the previous study (Chen et al., 2020), T241-T21 and T33-T241 were Two-Category assimilation contrasts; T33-T21 was a Single-Category assimilation contrast; and T315-T45 and T33-T45 were UnCategorised-Categorised assimilation contrasts.

3.1. Method

3.1.1. Participants

We tested 32 native speakers of Southern Vietnamese, divided into two memory load conditions with 16 participants each (Low: $M_{age} = 24.4$ years, $SD = 7.7$ years, 12 females, 4 males; High: $M_{age} = 27.2$ years, $SD = 12.8$ years, 13 females, 3 males). They completed the same discrimination and categorisation tasks with the same Thai stimuli as in Experiment 1.

3.1.2. Stimuli and procedure

These were the same as in Experiment 1. The only difference was that for the categorisation task, the Vietnamese participants were asked to categorise Thai tones using their five⁵ native tone categories. Stickers with the Vietnamese words in their native orthography were placed on the keys, i.e., “f”, “g”, “h”, “j”, “k”, in the same line on the keyboard, for their tone categorisation choices.

3.2. Results

3.2.1. Categorisation

To determine assimilation type of Thai tones with the present data, the same criteria were employed as Experiment 1, except that chance level for Vietnamese speakers was 20 % given that Southern Vietnamese has five tones. Missing data accounts for 0.58 % of the total data. We followed the same procedures as in Experiment 1 for determining Categorised or UnCategorised assimilation (see Appendix B, Table B.1,

Table B.2, and Table B.3 for details on the statistical tests). Three Thai tones were Categorised into native Vietnamese tone categories under both memory loads: T21 was Categorised to SV22, T241 was Categorised to SV44, and T315 was Categorised to SV214. Both T33 and T45 were UnCategorised under low memory load but Categorised to SV22 and SV21 respectively under high memory load (see Table 5).

We fitted the data with multinomial regression models as in Experiment 1. The full model was built with Vietnamese tone choices as a dependent measure, and memory load, talker variability, vowel variability and Thai tones (T21, T33, T45, T241, T315) as fixed factors. Four additional models were built each with one fixed factor subtracted and these models were compared to the full model to determine the effects of cognitive load. Both memory loads, $\chi^2(80) = 190.21$, $p < 0.001$, and Thai tone types, $\chi^2(128) = 7243.49$, $p < 0.001$, showed significant effects in Likelihood ratio tests. However, the talker variability and vowel variability effects were not significant.

3.2.2. Predictions for discrimination

First, we identified phonological assimilation and overlap types for the Vietnamese group based on the current perceptual assimilation results. The T33-T45 contrast resulted in UnCategorised-UnCategorised/Non-overlap assimilation under low memory load but Two-Category/Non-overlap under high memory load. The T241-T21 contrast was a Two-Category/Non-overlap assimilation under both memory loads. The T315-T45 contrast yielded UnCategorised-Categorised/Partial-overlap assimilation under low memory load but Two-Category/Non-overlap assimilation under high memory load. T33-T21 was UnCategorised-Categorised/Partial-overlap under low memory load but Single-Category/Non-overlap under high memory load, where the difference in ratings was not significant, $t(28) = -0.88$, $p = 0.38$. T33-T241 was UnCategorised-Categorised/Partial-overlap under low memory load but Two-Category/Partial-overlap under high memory load. Thus, the predictions for discrimination accuracy based solely on the assimilation type and phonological overlap are, from good to poor: under low memory load, Two-Category/Non-overlap {T241-T21} = UnCategorised-UnCategorised/Non-overlap {T33-T45} > UnCategorised-Categorised/Partial-overlap {T315-T45 = T33-T241 = T33-T21}; and under high memory load, Two-Category/Non-overlap {T241-T21 = T33-T45 = T315-T45} > Two-Category/Partial-overlap {T33-T241} = Single-Category/Non-overlap {T33-T21}.

As in Experiment 1, we calculated both the mean and relevant 95 % confidence intervals for the overlap scores and fit index difference scores (Appendix B, Table B.4), and used the confidence intervals to adjust predictions about discrimination due to differences in residual phonetic sensitivities. Under low memory load, there were no significant residual phonetic sensitivity differences between T33-T45 and T241-T21, nor among T315-T45, T33-T241 and T33-T21. Thus, our original predictions on discrimination performance remained unchanged. Under high memory load, the overlap scores but not the fit index difference scores showed differences: T33-T45 ($M_{Overlap_score} = 0.007$) had lower overlap scores than both T241-T21 ($M_{Overlap_score} = 0.204$) and T315-T45 ($M_{Overlap_score} = 0.312$) but the latter two contrasts did not differ. The predictions on discrimination performance were therefore

⁵ The standard dialect of Vietnamese (Northern dialect) has six tones but Chen et al. (2020) showed that Southern Vietnamese speakers produced both V214 (hỏi) and V415 (ngã) as a tone merger (no significant acoustic differences) consistent with previous reports (Brunelle, 2009).

Table 5

Assimilation of Thai tones into Vietnamese tone categories under low versus high memory loads. Categories significantly above chance (20 %) are in bold; "***" = Categorical tone. Assimilations: C = Categoricalised, U = UnCategoricalised. Ratings: 1 = Poor, 7 = perfect; mean ratings are displayed. "-" = no responses.

Memory load	Response	Thai stimulus									
		T45		T33		T21		T315		T241	
		%	rating	%	rating	%	rating	%	rating	%	rating
Low	SV44	6.3	3.5	43.5	5.4	4.3	3.6	0.2	1.0	82.7*	5.7
	SV22	2.2	2.4	51.8	5.3	88.7*	5.4	1.8	3.4	13.2	5.4
	SV35	24.6	4.7	1.1	2.2	0.5	4.5	7.6	5.8	1.8	3.5
	SV21	42.7	4.3	3.6	3.6	6.3	3.9	5.6	3.5	1.6	3.8
	SV214	24.1	4.3	-	-	0.2	1.0	84.8*	5.2	0.7	3.3
Assimilation		U _{clustered}		U _{clustered}		C		C		C	
High	SV44	0.4	1.0	38.1	5.0	3.4	3.1	0.2	1.0	81.1*	4.8
	SV22	0.4	2.5	61.6*	5.0	93.7*	4.7	-	-	16.6	4.5
	SV35	15.4	3.9	-	-	-	-	1.6	4.3	1.8	3.0
	SV21	60.9*	4.0	0.2	1.0	1.6	2.3	6.8	2.8	0.4	1.5
	SV214	22.8	3.0	-	-	1.3	2.5	91.4*	4.0	-	-
Assimilation		C		C		C		C		C	

changed from T33-T45 = T241-T21 = T315-T45, as predicted only based on assimilation and overlap types, to T33-T45 > T241-T21 = T315-T45. All other predictions listed earlier remain unchanged.

3.2.3. Discrimination

The d' scores for each participant were calculated separately for each tone pair as the dependent variable. Trials with time-outs were missing data, which accounted for 1.26 % of all the data across all participants in this task. We modelled the data using a Linear Mixed Effects Regression (LMER) model and calculated p values as in Experiment 1 (see Table 6, for the coefficients of the fixed effects and interactions see Table B.5 in the Appendix B).

The main effect of memory load was non-significant. However, there were significant main effects of talker variability, vowel variability, and Thai tone contrasts. The scores for the constant-talker blocks ($M = 3.15$, 95 % CIs [3.00, 3.30]) were significantly higher, suggesting better discrimination, than those of the variable-talker blocks ($M = 2.72$, 95 % CIs [2.56, 2.87]). Similarly, the scores for the constant-vowel blocks ($M = 3.20$, 95 % CIs [3.06, 3.35]) were significantly higher than

those of the variable-vowel blocks ($M = 2.67$, 95 % CIs [2.51, 2.82]).

To further examine Thai tone contrast main effects and test our predictions based on assimilation patterns, multiple comparisons with Tukey adjustments were conducted as in Experiment 1. All combinations of Thai tone contrasts were significantly different from each other, except for the comparison between T33-T21 and T33-T241 (see Appendix B, Table B.5). The results reflect the following overall pattern: T33-T45 ($M = 4.08$, 95 % CIs [3.86, 4.30]) > T241-T21 ($M = 3.64$, 95 % CIs [3.46, 3.82]) > T315-T45 ($M = 2.69$, 95 % CIs [2.47, 2.92]) > T33-T21 ($M = 2.28$, 95 % CIs [2.05, 2.51]) = T33-T241 ($M = 1.98$, 95 % CIs [1.83, 2.13]).

In addition to the main effects, multiple comparisons were conducted with Tukey adjustments to break down the talker variability $\times$ vowel variability interaction (see Fig. 4, Panel A). In the constant-talker condition, the constant-vowel blocks ($M = 3.59$, 95 % CIs [3.41, 3.78]) were better discriminated than the variable-vowel blocks ($M = 2.71$, 95 % CIs [2.49, 2.93]), $\beta = 0.88$, $SE = 0.11$, $t(107) = 7.91$, $p < 0.001$. Within the constant-vowel blocks, the constant-talker blocks ($M = 3.59$) were better discriminated than the variable-talker

Table 6

Main effects and interactions for Thai tone contrast discrimination by Vietnamese listeners. "***" indicates significant.

Cognitive factors	F	df		p
Memory load	0.72	1	30	0.40
Talker variability	30.77	1	30	< 0.01*
Vowel variability	44.82	1	30	< 0.01*
Tone contrasts	105.92	4	510	< 0.01*
Memory load $\times$ Talker variability	0.09	1	30	0.77
Memory load $\times$ Vowel variability	6.40	1	30	0.02*
Talker variability $\times$ Vowel variability	19.79	1	510	< 0.01*
Memory load $\times$ Tone contrasts	2.39	4	510	0.05*
Talker variability $\times$ Tone contrasts	0.82	4	510	0.51
Vowel variability $\times$ Tone contrasts	2.65	4	510	0.03*
Memory load $\times$ Talker variability $\times$ Vowel variability	0.19	1	510	0.67
Memory load $\times$ Talker variability $\times$ Tone contrasts	0.30	4	510	0.88
Memory load $\times$ Vowel variability $\times$ Tone contrasts	0.56	4	510	0.69
Talker variability $\times$ Vowel variability $\times$ Tone contrasts	0.72	4	510	0.58
Memory load $\times$ Talker variability $\times$ Vowel variability $\times$ Tone contrasts	0.22	4	510	0.93

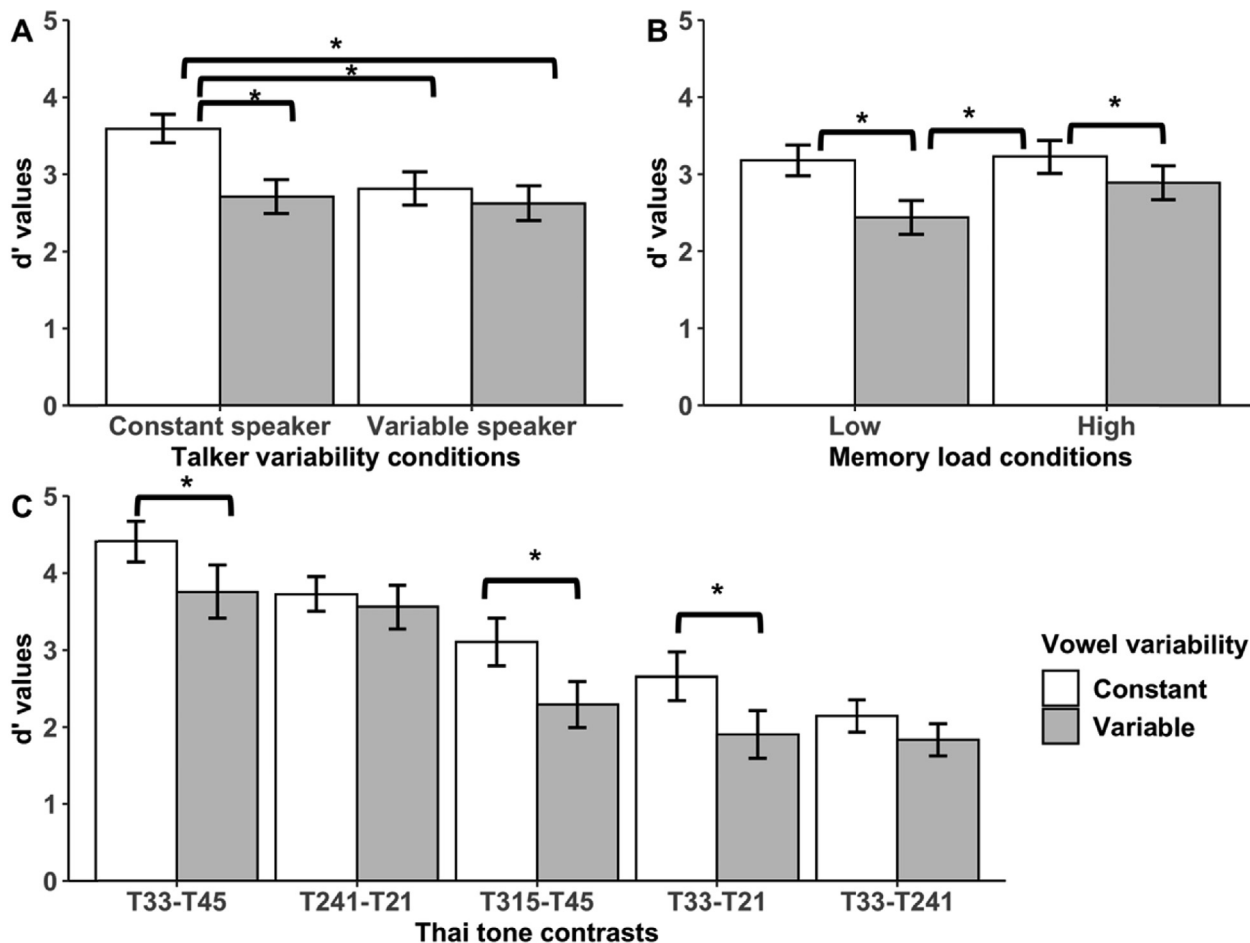


Fig. 4. Interactions between vowel variability and talker variability (panel A), vowel variability and memory load (panel B), vowel variability and tone contrasts (panel C). d' plotted along the y axis was calculated using a differencing rule for AX tasks (Macmillan & Creelman, 2005). Error bars indicate 95% confidence intervals around the mean.

blocks ($M = 2.81$, 95 % CIs [2.60, 3.03]), $\beta = 0.78$, $SE = 0.11$, $t(113) = 7.07$, $p < 0.001$. In addition, the constant-talker + vowel block had significantly higher scores ($M = 3.59$) than the variable-talker + vowel block, ($M = 2.62$, 95 % CIs [2.40, 2.85]), $\beta = 0.96$, $SE = 0.11$, $t(60) = 8.67$, $p < 0.001$.

Similarly, we conducted multiple comparisons with Tukey adjustments on the memory load $\times$ vowel variability interaction (see Fig. 4, Panel B). Under low memory load, the constant-vowel conditions ($M = 3.18$, 95 % CIs [2.98, 3.38]) were better discriminated than the variable-vowel conditions ($M = 2.44$, 95 % CIs [2.22, 2.66]), $\beta = 0.74$, $SE = 0.114$, $t(30) = 6.52$, $p < 0.001$. Similarly, when the memory load was high, the constant-vowel conditions ($M = 3.23$, 95 % CIs [3.01, 3.44]) were also better discriminated than the variable-vowel conditions ($M = 2.89$, 95 % CIs [2.67, 3.11]), $\beta = 0.33$, $SE = 0.11$, $t(30) = 2.95$, $p = 0.03$. In addition, the constant-vowel blocks under high memory load ($M = 3.23$) were better discriminated than the variable-vowel blocks under low memory load ($M = 2.44$), $\beta = 0.79$, $SE = 0.20$, $t(41) = 4.02$, $p = 0.001$.

We also conducted multiple comparisons with Tukey adjustments to break down the interaction between vowel variability and tone contrasts (see Fig. 4, Panel C). The constant-vowel blocks were discriminated better than the variable-vowel blocks for: **T33-T45**, $\beta = 0.65$, $SE = 0.18$, $t(370) = 3.74$,

$p < 0.01$; **T315-T45**, $\beta = 0.81$, $SE = 0.18$, $t(370) = 4.61$, $p < 0.001$; and **T33-T21**, $\beta = 0.75$, $SE = 0.18$, $t(370) = 4.3$, $p < 0.01$.

Lastly, we ran multiple comparisons with Tukey adjustments to break down the just significant ($p = 0.05$) interaction between memory load and tone contrasts. However, there were no significant differences when comparing the same tone contrast between the two memory loads.

3.3. Discussion

In Experiment 2 as in Experiment 1, categorisation patterns were modulated by memory load. T45 and T33 were UnCategorised assimilations under low memory load but became Categorised under high memory load, reflecting stronger native phonological constraints in the latter case, consistent with our reasoning that Categorised assimilations reflect stronger native phonological constraints than UnCategorised assimilations. On the other hand, the high percent choice of Categorised native responses reflects low phonetic sensitivity to deviations of non-native tones from native categories. The percent choice of Categorised assimilations for T21 and T315 increased under high memory load, suggesting reduced resid-

ual sensitivity to phonetic deviations between native and non-native categories.

Similar to Experiment 1, PAM-driven predictions based on Vietnamese assimilation patterns with consideration of phonological overlap types and phonetic overlap scores were supported by the discrimination results. The T33-T21 contrast was a Single-Category contrast assimilation, as predicted, under high memory load and was discriminated more poorly than T33-T45, T241-T21 and T315-T45, which were all assimilated as Two-Category/Non-overlap under high memory load. Phonological overlap reduced discrimination accuracy of non-native contrasts. The T241-T21 contrast, Two-Category/Non-overlap, under low memory load, was discriminated more accurately than T315-T45, UnCategorised-Categorised/Partial-overlap under low memory load. Similarly, the T33-T241 contrast was an UnCategorised-Categorised/Partial-overlap assimilation under low memory load but Two-Category/Partial-overlap under high memory load. In both cases, phonological overlap in native response categories reduced its discrimination accuracy to be as poor as T33-T21. Phonetic sensitivity as indicated by overlap scores, but not fit-index difference scores, successfully predicted more accurate discrimination of T33-T45 than T241-T21, both being Two-Category/Non-overlap under high memory load.

As in Experiment 1, categorisation but not discrimination was affected by memory load manipulations. We interpret this to indicate that delaying categorisation responses biased listeners toward phonological processing under high memory load, and consequently the results were more constrained by native phonological factors under high than under low memory load. In discrimination, however, we again found no main effects of memory load, supporting the idea that tones persist longer in working memory than consonants, which have shown effects of memory load on categorisation.

Moreover, as in Experiment 1, talker variability and vowel variability affected discrimination but not categorisation. We speculated that categorisation elicited phonological processing as it required listeners to assimilate non-native phones into native phonological categories and listeners may rely on mechanisms used in their native language to maintain phonological constancy, i.e., the ability to keep word identity and thus the abstract phonological category of lexical tone intact, and accommodate to linguistically irrelevant variability. On the other hand, discrimination only requires listeners to compare two non-native phones which should result in more of a focus on phonetic details. In variable talker/vowel blocks, nevertheless, we infer that listeners tended to use a more phonological mode of perception to achieve phonological constancy for the tones, and thus were less sensitive to phonetic details and more constrained by native tone phonology. Discrimination in this case was less accurate than in constant talker/vowel blocks where we infer that listeners shifted to a more phonetic mode of perception and were more sensitive to the specific, concrete f0 properties.

3.4. Cross-language comparison of Thai tone contrast discrimination

The analysis thus far has revealed how assimilation patterns modulated discrimination of tone contrasts within each language group. In addition, the discrimination of each tone

contrast across two language groups was compared to see whether cross-linguistic PAM-based predictions about differences between the two language groups are also supported. If so, the same tone contrast assimilated differently by the two groups should be reflected in different discrimination performance.

To this end, we ran a Linear Mixed Effect Regression (LMER) model with d' as the dependent variable, entering native languages, memory load, talker variability, vowel variability and tone contrasts (T241-T21, T33-T21, T315-T45, T33-T241, T33-T45) as fixed factors, and participant as a random factor including random slopes for within-subject fixed factors, i.e., talker and vowel variability conditions (see Appendix B, Table B.7 and Table B.8 for full statistical details). There were main effects of talker variability, $F(1, 60) = 71.07$, $p < 0.001$, vowel variability, $F(1, 60) = 101.33$, $p < 0.001$, and Thai tone contrasts, $F(4, 1020) = 198.98$, $p < 0.001$, and two interactions, talker variability $\times$ vowel variability, $F(1, 1020) = 39.80$, $p < 0.001$, and language $\times$ Thai tone contrasts, $F(4, 1020) = 32.80$, $p < 0.001$, but no main effect of language group, indicating that neither language group had an overall advantage in discriminating overall across the five Thai tone contrasts, but that their performance diverged for at least some tone contrasts.

We conducted multiple comparisons with Tukey adjustments to explore the interaction between language groups and Thai stimuli (see Appendix B, Table B.9 for the full comparisons). The discrimination of two contrasts, T241-T21 and T33-T21, showed significant differences between the two language groups, while discrimination of the other three contrasts T315-T45, T33-T241, and T33-T45 did not differ between groups. Consistent with PAM predictions, the T241-T21 contrast was better discriminated by Vietnamese listeners, who assimilated it as Two-Category/Non-overlap, $M = 3.64$, 95 % CIs [3.46, 3.82], than by Mandarin listeners, who assimilated it as Category-Goodness/Non-overlap, $M = 3.05$, 95 % CIs [2.85, 3.26], $\beta = -0.59$, $SE = 0.17$, $t(153) = -3.40$, $p = 0.03$. Conversely, the T33-T21 contrast was discriminated much better by Mandarin listeners, who assimilated it as Two-Category/Non-overlap, $M = 3.42$, 95 % CIs [3.21, 3.63], than by Vietnamese listeners, who assimilated it as UnCategorised-Categorised/Partial-overlap (low memory load) or Single-Category/Non-overlap (high memory load), $M = 2.28$, 95 % CIs [2.05, 2.51], $\beta = 1.14$, $SE = 0.17$, $t(153) = 6.58$, $p < 0.01$.

4. General discussion

The aims of the two experiments in the present study were to examine native phonological and phonetic influences on the perceptual assimilation and discrimination of non-native lexical tones and to determine whether and how memory load, talker and vowel variability shift listeners between a phonological versus phonetic mode.

4.1. Native languages influence on non-native tone perception

Native language tone systems shaped the perception of non-native tones for both language groups as predicted by PAM principles. When two non-native tones were categorised to the same native tone category, i.e., Single-Category or

Category-Goodness assimilations, discrimination was less accurate than when two non-native tones were neatly categorised into two native tone categories, i.e., Two-Category assimilation.

In addition, perceived phonological overlap, i.e., complete, partial and no overlap, in listeners' assimilations of non-native tone contrasts affected their discrimination even for the same type of contrast assimilation. Contrasts with greater phonological overlap in categorisation were discriminated more poorly than those with a smaller degree of overlap within each of the three assimilation types: Two-Category, UnCategorised-Categorised and UnCategorised-UnCategorised. The effect of phonological overlap in reducing discrimination accuracy can be large. For example, T33-T241 was among the worst discriminated contrasts for both groups even though it was assimilated as Two-Category contrast, when assimilations of it showed partial overlap under high memory load.

Phonetic effects can refine some predictions on discrimination of the same/similar assimilation-overlap type contrast. For example, T33-T45 and T33-T21 were both Two-Category/Non-overlap assimilations for Mandarin listeners. Nonetheless, both overlap scores and fit index difference scores successfully predicted T33-T45 to be better discriminated than T33-T21, suggesting that residual phonetic sensitivity to non-native phones from native categories thus should be considered to make more precise PAM predictions of contrasts of the same assimilation-overlap type.

Comparing discrimination performance across two language groups provides strong evidence for native tone system influences. For the same non-native tone contrast, different assimilation patterns between the two language groups led to different levels of discrimination accuracy, consistent with PAM principles. For example, T21 was assimilated into the high falling tone M51 by Mandarin listeners who have no low level tone but was assimilated to the low level tone SV22 by Vietnamese listeners who have a height contrast between two level tones. Additionally, mid-level tone T33 was assimilated as high level tone M55 by Mandarin listeners, but as low level tone SV22 by Vietnamese listeners under high memory load and split between high-level tone SV44 and SV22 under low memory load. Thus, the T33-T21 contrast forms a Two-Category/Non-overlap assimilation for Mandarin listeners but an UnCategorised-Categorised/Partial-overlap or Single-Category/Non-overlap assimilation for Vietnamese listeners. T33-T21 was better discriminated by Mandarin than Vietnamese listeners in line with PAM predictions. This type of comparison could not be assessed in previous studies that compared non-native listeners of non-tone languages with listeners of a single tone language, because tone and non-tone language listeners assimilate non-native tones at different levels of their native phonological systems (Best, 2019).

4.2. Memory load effects on non-native perception

Although we kept the memory load manipulation at identical lengths (500 ms versus 2000 ms) across the categorisation and discrimination tasks, manipulating the temporal course of the judgements in the tasks exerted different effects on categorisation and discrimination. In the categorisation task, after

hearing the stimulus, listeners could start processing immediately and some could even make a decision before the signal to respond, which is different from the discrimination task where listeners were unable to make the decision until they heard the second stimulus. However, despite the possibility that listeners could make their decision before the "respond" signal in the categorisation task, the memory manipulation clearly did influence categorisation, as we found significant memory load effects in both language groups on that task. We speculate that the delayed RI in the categorisation task not only increased memory load but also lengthened the processing time for the target stimulus so that listeners were more likely to engage in higher level phonological processing, i.e., a phonological mode, and thereby based their decisions on perceived abstract pitch contours. We further speculate that short RI, on the other hand, pushed listeners to base their categorisation choice more on specific, concrete f0 properties, i.e., a phonetic mode, resulting in insufficient processing time to complete phonological assimilation before making their choices.

Future research could aim to clearly interrupt processing and prevent listeners from making their categorisation decision prior to the signal to respond by adding an intervening task, such as counting digits. The reason we did not include an intervening task in the current study was that we needed the categorisation task to be comparable to the discrimination task, which did not use an intervening task. Moreover, as the memory load manipulation did yield significant effects in the categorisation task, it clearly was effective as expected in our categorisation task.

On the other hand, in the discrimination task, listeners need to hold the first stimulus in memory to compare it with the second stimulus and decide whether the two were the same or different. Longer ISI did not affect discrimination, compatible with previous findings on non-native perception of Mandarin and Cantonese tones (Lee et al., 1996). This lends support to the cue-duration hypothesis (Fujisaki & Kawashima, 1970), according to which phonetic details of longer-duration speech segments such as vowels, and as we argue tones, are better retained in memory. In other words, the long ISI may not have been long enough to yield sufficient fading of phonetic details to reduce performance relative to the short ISI. Future research can test this hypothesis by using a longer ISI, such as 2500 ms (Asano, 2018), for the high memory load condition.

Our memory load manipulation affected assimilation but not discrimination. Thus, our two sets of predictions based on assimilation results under low versus high memory load did not align with discrimination performance, which did not differ between the two memory load conditions. For example, for Vietnamese listeners, better discrimination of T241-T21 than T315-T45 was predicted by assimilation types under low memory load, i.e., Two-Category/Non-overlap > UnCategorised-Categorised/Partial-overlap, but assimilation patterns under high memory load predicted equivalent discrimination performance. On the other hand, assimilation patterns under high memory load successfully predicted better discrimination of the contrast T315-T45 Two-Category/Non-overlap than T33-T241 Two-Category/Partial-overlap but assimilation patterns under low memory load predicted equally good discrimination. Nevertheless, when the two sets of assimilation-based predictions were combined, they did account for discrimination variations across

different contrasts. Given that assimilation patterns indicated varying degrees of native language phonological and phonetic contributions under high or low memory load, when they successfully predicted discrimination accuracy, they did reveal the amount of native phonological and phonetic contributions to discrimination. Future research should consider the effects of memory load in modulating assimilation to provide more accurate predictions of discrimination.

4.3. Talker and phonetic context variability effects on non-native tone perception

While stimulus variability, i.e., talker and vowel variability, affected discrimination performance, categorisation was largely immune to its effects. We propose that the differing stimulus variability effects observed in categorisation and discrimination reflect different processes subserving the two tasks, as has been argued in previous studies (Antoniou, Tyler, & Best, 2012, 2013). Categorisation tasks require phonological judgements, i.e., categorising non-native tones into native categories. Thus, naïve listeners are more likely to use their native language phonological categories as well as mechanisms used to maintain phonological constancy in the native language when processing the non-native stimulus. In other words, perceivers assimilate not only non-native tones to native categories but also assimilate the phonetic properties of unfamiliar talkers to their native phonological categories (Best, 2015). Since the listeners in the present study were all native tone language speakers, their phonological representation of native tones should include variations induced by vowel contexts and talkers.

Conversely, in discrimination, listeners simply decide whether the tones in the stimulus pairs were the same or not. Thus, listeners base their decisions on phonetic information, i.e., f_0 properties, and are more likely to be affected by talker and/or vowel variations. Compatibly, listeners generally did discriminate the same contrast better in constant talker and vowel than in variable talker and vowel blocks. This supports our hypothesis that listeners use more of a phonetic mode in constant talker and vowel blocks, where phonetic details were more reliable than those in variable talker and vowels blocks.

There are two additional differences between categorisation and discrimination tasks regarding talker and vowel variability. First, the effects of talker and vowel variability in the categorisation task were cross-trial effects because there was only one tone in each trial. In contrast, in the discrimination task the talker and vowel variability in each block was within-trial, i.e., the A and the X had either the same or different talker(s)/vowel(s). In other words, variability was generally higher in the discrimination task relative to the categorisation task in the variable talker and/or vowel blocks. Additionally, in the discrimination task, two perceptual dimensions were involved and interfered with each other, especially in the variable-vowel condition. Listeners were required to attend selectively to the tone contour/height while simultaneously ignoring vowel and/or talker differences between the matching tone stimuli. We speculate that this required more attentional control in high variability conditions and biased listeners to use a phonological mode of perception. Further research would be needed to confirm this interpretation.

5. Conclusion

Discrimination of non-native tone contrasts is significantly influenced by how non-native tones are assimilated into native tone categories. A novel contribution of our research is that both perceived phonological overlap types and phonetic overlap indices can account for variations in discrimination of non-native contrasts of the same assimilation types, suggesting that both phonological and phonetic factors should be considered when predicting non-native discrimination. Moreover, delaying responses in categorisation, i.e., under high memory load, leads listeners to process the stimuli in a more phonological mode, resulting in more phonologically-based assimilations, whereas allowing listeners to respond more quickly, as under low memory load, leads them to base their choices more on the phonetic properties of the stimuli. On the other hand, listeners were not affected by talker or vowel variability in categorisation, which suggests that they used their native language phonological categories to accommodate these phonologically-irrelevant variations. Unlike categorisation, high stimulus variability in the discrimination task biased listeners to a phonological mode of perception whereas memory load did not have any effect. Although native language influences remain a primary factor as predicted by PAM, memory load and stimulus variability biased categorisation and discrimination differently toward phonological versus phonetic mode, consistent with the principles of ASP. The current findings thus have substantive implications for theories of tone perception as well as for second language lexical tone learning. Native language phonological and phonetic factors as modulated by phonological and phonetic modes in categorisation and discrimination should be considered when researching and teaching non-native tone perception and learning.

CRediT authorship contribution statement

Juqiang Chen: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization.
Mark Antoniou: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Supervision.
Catherine T. Best: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that there is no conflict of interest.

Acknowledgements

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

Appendix A and B. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.wocn.2022.101199>.

References

- Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics*, 40(4), 582–594. <https://doi.org/10.1016/j.wocn.2012.05.005>.
- Antoniou, M., Best, C. T., & Tyler, M. D. (2013). Focusing the lens of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *The Journal of the Acoustical Society of America*, 133(4), 2397–2411. <https://doi.org/10.1121/1.4792358>.
- Asano, Y. (2018). Discriminating non-native segmental length contrasts under increased task demands. *Language and Speech*, 61(3), 409–429. <https://doi.org/10.1177/0023830917731907>.
- 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. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1).
- Baddeley, A. (2010). Working memory. *Current Biology*, 20(4), R136–R140. <https://doi.org/10.1016/j.cub.2009.12.014>.
- 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. <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) <https://doi.org/10.18637/jss.v067.i01>.
- Best, C., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20(3), 305–330.
- 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. <https://doi.org/10.1111/j.1467-9280.2009.02327.x>.
- Best, C., Avesani, C., Tyler, M. D., & Vayra, M. (2019). PAM revisits the Articulatory Organ Hypothesis: Italians' perception of English anterior and Nu-Chah-Nulth posterior voiceless fricatives. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A Sound Approach to Language Matters* (pp. 13–40). In Honor of Ocke-Schwen Bohn.
- 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.
- 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. <https://doi.org/10.3389/fpsyg.2019.00364>.
- Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, 37(1), 79–96. <https://doi.org/10.1016/j.wocn.2008.09.003>.
- Burnham, D., & Jones, C. (2002). Categorical perception of lexical tone by tonal and non-tonal language speakers. In *Proceedings of the Ninth Australian International Conference on Speech Science and Technology* (pp. 515–520).
- Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. In *Proceeding of Fourth International Conference on Spoken Language Processing*. <https://doi.org/10.1109/ICSLP.1996.607325>.
- 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>.
- Chang, Y. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America*, 141(2), EL120–EL126. <https://doi.org/10.1121/1.4976037>.
- Chao, Y. R. (1930). A system of tone-letters. *Le Maître Phonétique*, 45, 24–27.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press.
- Chen, J., Best, C. T., & Antoniou, M. (2020). Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners. *Journal of Phonetics*, 83. <https://doi.org/10.1016/j.wocn.2020.101013>.
- Duanmu, S. (1990). *A formal study of syllable, tone, stress and domain in Chinese languages*. Massachusetts Institute of Technology.
- Duanmu, S. (1994). Against Contour Tone Units. *Linguistic Inquiry*, 25(4), 555–608.
- Escudero, P., & Vasiliev, P. (2011). Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels. *The Journal of the Acoustical Society of America*, 130(5), EL277–EL283. <https://doi.org/10.1121/1.3632043>.
- Escudero, P., & Williams, D. (2011). Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *The Journal of the Acoustical Society of America*, 129(1), EL1–EL7. <https://doi.org/10.1121/1.3525042>.
- Escudero, P. (2005). Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization. LOT.
- Ewan, W. G. (1975). Explaining the intrinsic pitch of vowels. *The Journal of the Acoustical Society of America*, 58(S1), S40. <https://doi.org/10.1121/1.2002115>.
- 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. <https://doi.org/10.1016/j.wocn.2018.05.003>.
- Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26(1), 1–34. <https://doi.org/10.1017/S0272263104026117>.
- 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.
- Fox, J., & Weisberg, S. (2019). An R companion to applied regression (3rd ed.). Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Fujisaki, H., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. (pp. 29, 207–214) [Annual Report of the Engineering Research Institute]. University of Tokyo.
- Gandour, J. T. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 41–76). Academic Press.
- 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. <https://doi.org/10.1121/1.4783674>.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R Package. *Journal of Statistical Software*, 33(2). <https://doi.org/10.18637/jss.v033.i02>.
- 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.
- Hallé, P. A., Best, C., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, 27(3), 281–306. <https://doi.org/10.1006/pho.1999.0097>.
- Hao, Y.-C. (2017). Second language perception of Mandarin vowels and tones. *Language and Speech*, 61(1), 135–152. <https://doi.org/10.1177/0023830917717759>.
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers*, 27(1), 46–51. <https://doi.org/10.3758/BF03203619>.
- Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "Perceptual Magnet Effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). York Press.
- Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25(5), 527–542. <https://doi.org/10.1007/BF01758181>.
- Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, 69(1), 1–33. <https://doi.org/10.18637/jss.v069.i01>.
- Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults' French vowel learning. *The Journal of the Acoustical Society of America*, 126(5), 2670–2682. <https://doi.org/10.1121/1.3224715>.
- Lin, Y.-H. (1989). *Autosegmental treatment of segmental processes in Chinese phonology*. University of Texas at Austin [Unpublished Doctoral dissertation].
- Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, 49(2B), 606–608. <https://doi.org/10.1121/1.1912396>.
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4), 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y>.
- MacKain, K. S., Best, C., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2(4), 369–390. <https://doi.org/10.1017/S01421716400009796>.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Psychology press.
- Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2), 391–409. <https://doi.org/10.1037/0096-1523.33.2.391>.
- Miller, J. (1996). The sampling distribution of d'. *Perception & Psychophysics*, 58(1), 65–72. <https://doi.org/10.3758/BF03205476>.
- Mitterer, H., & Mattys, S. L. (2017). How does cognitive load influence speech perception? An encoding hypothesis. *Attention, Perception, & Psychophysics*, 79(1), 344–351. <https://doi.org/10.3758/s13414-016-1195-3>.
- Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18(5), 331–340. <https://doi.org/10.3758/BF03211209>.
- Mullenix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47(4), 379–390. <https://doi.org/10.3758/BF03210878>.
- Mullenix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, 85(1), 365–378. <https://doi.org/10.1121/1.397688>.
- Nhàn, N. T. (1984). The syllabeme and patterns of word formation in Vietnamese [PhD Dissertation]. 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.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253–260. <https://doi.org/10.3758/BF03214136>.
- 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. <https://doi.org/10.3758/s13414-014-0791-3>.
- Shaw, J. A., & Tyler, M. D. (2020). Effects of vowel coproduction on the timecourse of tone recognition. *The Journal of the Acoustical Society of America*, 147(4), 2511–2524. <https://doi.org/10.1121/10.0001103>.
- Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3(3), 243–261. Scopus. <https://doi.org/10.1017/S0142716400001417>.
- So, C. K., & Best, C. (2014). Phonetic influences on English and French listeners' assimilation of mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, 36(2), 195–221. <https://doi.org/10.1017/S0272263114000047>.
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39(4), 456–466. <https://doi.org/10.1016/j.wocn.2010.09.001>.
- Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (2019). Acquisition of weak syllables in tonal languages: Acoustic evidence from neutral tone in Mandarin Chinese. *Journal of Child Language*, 46(1), 24–50. <https://doi.org/10.1017/S0305000918000296>.
- Tsukada, K., & Kondo, M. (2018). The perception of Mandarin lexical tones by native speakers of Burmese. *Language and Speech*, 62(4), 625–640. <https://doi.org/10.1177/0023830918806550>.
- Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, 71(1), 4–21. <https://doi.org/10.1159/000356237>.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* ((4th ed.)). Springer.
- Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, 97(1), 144–160. <https://doi.org/10.1111/j.1540-4781.2013.01386.x>.
- Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, 54(4), 681–712. <https://doi.org/10.1111/j.1467-9922.2004.00283.x>.
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37(1), 35–44. <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.
- Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23(3), 349–366. [https://doi.org/10.1016/S0095-4470\(95\)80165-0](https://doi.org/10.1016/S0095-4470(95)80165-0).
- Wiener, S., & Lee, C.-Y. (2020). Multi-talker speech promotes greater knowledge-based spoken Mandarin word recognition in first and second language listeners. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00214>.
- 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.
- Wu, X., Munro, M. J., & Wang, Y. (2014). Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics*, 46, 86–100. <https://doi.org/10.1016/j.wocn.2014.06.005>.
- Wu, M., Bundgaard-Nielsen, R. L., Baker, B., Best, C., & Fletcher, J. (2015). Perception of Cantonese tones by Mandarin speakers. *Proceedings of ICPHS 2015*.
- Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics*, 52(4), 376–392. <https://doi.org/10.3758/BF03206698>.
- Yu, Y. H., Shafer, V. L., & Sussman, E. S. (2017). Neurophysiological and behavioral responses of Mandarin lexical tone processing. *Frontiers in Neuroscience*, 11. <https://doi.org/10.3389/fnins.2017.00095>.
- Zheng, Q. (2014). Effects of vowels on Mandarin tone categorical perception. *Acta Psychologica Sinica*, 46(9), 1223. <https://doi.org/10.3724/SP.J.1041.2014.01223>.