



Research Article

Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners

Juqiang Chen^{a,*}, Catherine T. Best^{a,b,*}, Mark Antoniou^a^a Western Sydney University, The MARCS Institute for Brain Behaviour and Development, Locked Bag 1797, Penrith, NSW 2751, Australia^b Haskins Laboratories, New Haven CT, USA

ARTICLE INFO

Article history:

Received 5 July 2019

Received in revised form 29 August 2020

Accepted 5 September 2020

Keywords:

Perceptual assimilation

Non-native lexical tones

Thai

Mandarin

Vietnamese

ABSTRACT

A cross tone-language perceptual assimilation study investigated native categorisations and goodness ratings of non-native Thai tones by Thai-naïve listeners differing in their native tone systems: Mandarin, Northern Vietnamese and Southern Vietnamese. We derived hypotheses from the Perceptual Assimilation Model (PAM: Best, 1995), which considers both native phonological and phonetic influences on perceptual assimilation of non-native speech contrasts. Mandarin listeners reliably categorised the Thai mid level tone to their single native level tone category, reflecting a native phonological effect, but they also showed high residual phonetic sensitivity to differences between the non-native tone and the native tone it was assimilated to, indicated by low category-goodness ratings. Native phonological and phonetic differences in tones of the two Vietnamese dialects also affected perceptual assimilation of the Thai high level and rising tones. In addition, categorisation responses were faster overall for Categorised than Uncategorised assimilations, revealing the processing cost of perceptual uncertainty due to phonological competition among and/or phonetic discrepancies from multiple native categories. This indicates, furthermore, a more focused and thus stronger native phonological contribution for Categorised than Uncategorised assimilations. PAM principles thus extend to non-native tone assimilations and indicate the importance of both native phonological and phonetic contributions to non-native speech perception.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

People form native phonological and phonetic categories through their experience hearing and speaking that language. This facilitates perception of the native language, but it reduces sensitivity to non-native phonemes/phones. Adults have difficulties in categorising and discriminating non-native consonants and vowels (Polka, 1992, 1995; Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001). However, the degree of difficulty can vary greatly, depending on the phonological and phonetic similarities and differences between the listeners' native (L1) language and the non-native language. Naïve listeners are unable to discern which non-native phonetic distinctions are phonologically contrastive and thus cannot distinguish between phonological and phonetic levels in the non-native stimuli (Best & Tyler, 2007). But they can access both phonological distinctions and phonetic details in

their own language as a reference framework for categorising unfamiliar non-native contrasts. Thus, native perceptual constraints can operate at either an abstract phonological level or a lower phonetic level in perception of non-native speech, or a combination of the two, to affect naïve listeners' responses to non-native phones. So far most studies on this issue have only examined consonants (Best & Strange, 1992; Bohn & Best, 2012; Hallé, Best, & Levitt, 1999) and vowels (Faris, Best, & Tyler, 2018; Tyler, Best, Faber, & Levitt, 2014).

More than 60% of the languages in the world are tone languages, in which pitch variations change the lexical meanings of words (Yip, 2002). However, relatively little is known about how native phonological and phonetic factors interplay in perceptual assimilation of non-native lexical tones. This question has not been resolved by previous cross-language assimilation studies with listeners of non-tone languages, such as English or French (So & Best, 2014). Such comparisons can reveal how non-tone language listeners perceive a type of segmental-level phonological element that does not exist in their phonological systems, unlike consonants and vowels which exist in

* Corresponding authors.

E-mail addresses: J.Chen2@westernsydney.edu.au (J. Chen), C.Best@westernsydney.edu.au (C.T. Best), m.antoniou@westernsydney.edu.au (M. Antoniou).

all languages (Best, 2019). In other words, non-tone language listeners may assimilate non-native tones into their native intonation categories, i.e. into prosodic, rather than segmental categories. However, that type of assimilation is likely to differ qualitatively from assimilation to a corresponding native tone category by listeners of other tone languages (Best, 2019). Since there are no functional lexical tone categories at the segmental level in the L1s of non-tone language listeners, studies of their perception of non-native tones cannot reveal how L1 segmental-tier phonological and phonetic factors affect non-native lexical tone perception.

Only a limited number of studies have actually tested non-native tone perceptual assimilation by tone language listeners, and they show inconsistent assimilation patterns (Reid et al., 2015; Wu, Munro, & Wang, 2014). The variations in perceptual assimilation reflect the complexity of tone perception in general and entail further investigation of the phonological versus phonetic basis of non-native tone assimilation, which have not been systematically examined in previous studies of non-native tone perception. It is important, as well, to include listeners of additional tone languages beyond Mandarin and Cantonese which are both spoken in China. The current study addresses this issue by examining perceptual assimilation of Thai tones by three listener groups, Mandarin, Northern Vietnamese and Southern Vietnamese. The phonological and phonetic differences and similarities among the target language and the listener languages provide an ideal ground for testing native tone language influences from both the phonological and phonetic levels of characterising lexical tones.

1.1. Phonological and phonetic influences on cross-language assimilation

To account for L1 influences on non-native speech perception, several theoretical models have been proposed: the Perceptual Assimilation Model (PAM, Best, 1995), the Native Language Magnet model (NLM, Kuhl & Iverson, 1995), the Speech Learning Model (SLM, Flege, 1995) and the Second Language Linguistic Perception model (L2LP: e.g., Escudero, 2005; Escudero & Vasiliev, 2011; Escudero & Williams, 2011). The focus of the current study is on adult tone-language listeners' perceptual assimilation of naturally produced non-native tones. NLM has its focus on within-category discrimination of synthetic vowels and consonants, primarily by infants. SLM has concentrated more on second language learners and their second language speech production, rather than primarily addressing perceptual assimilation (or perceived similarity) by naïve listeners, the interest of the current study. L2LP has focused mainly on second language learners' perception of vowels with predictions based only on acoustic similarities, not considering abstract phonological features. Thus, these models are not ideal to be employed as the main theoretical framework for exploring native language phonological and phonetic influences on non-native perceptual assimilation of lexical tones.

PAM is the model best suited to address the questions of the current study on non-native lexical tone assimilations by naïve adult listeners of other tone languages because it considers native language influence at both phonological and phonetic levels. At the phonetic level, PAM assumes that listeners

perceive the articulatory-phonetic information carried in speech, positing that naïve adult listeners perceptually assimilate non-native phones into native categories based on their perceived articulatory-phonetic (articulatory gestures) similarities to native phonemes. Articulatory gestures for lexical tones would involve laryngeal movements to raise and lower pitch (cricothyroid and arytenoid muscles) and possibly also to raise and lower the larynx itself, i.e., external muscles in the trachea. Laryngeal gestures may also result in voice quality changes such as breathiness and creakiness (Brunelle, Nguyễn, & Nguyễn, 2010; Erickson & Abramson, 2013; Erickson, 1976; Erickson, Liberman, & Niimi, 1976; Nguyen & Edmondson, 1998; Sagart, Hallé, de Boysson-Bardies, & Arabia-Guidet, 1986). However, extrapolating from PAM principles at the phonological level, we assume here that tone features are abstractions from articulatory-phonetic information. Therefore, in the present study we do not describe tone features in articulatory terms but rather as more abstract specifications. PAM will be employed as the main theoretical framework while NLM and SLM will also be compared and incorporated in the discussion of results where relevant.

Within the PAM framework, depending on phonological and phonetic similarities between a given non-native phone and the listener's native phonemes, they will perceptually assimilate it in one of the following ways: (i) as an exemplar of a native phoneme (Categorised) but will also perceive its phonetic goodness of fit to that native category (ranging from excellent to poor fit); or (ii) as unlike any single native phoneme but within the native phonological space, i.e., perceived as speech (Uncategorised), or (iii) as a nonspeech sound (Non-Assimilable). Faris, Best, and Tyler (2016, 2018) have extended PAM principles to account for three different ways in which an Uncategorised non-native phone can be perceived: (1) a *focalised* response in which the non-native phone is assimilated as primarily similar to a single L1 category but choices of that native phoneme fall below the defined categorisation threshold; (2) a *clustered* response in which the non-native phone is assimilated below threshold to a small set of L1 categories; or (3) a *dispersed* response in which the assimilation of a non-native phone category is spread across many L1 categories (all below chance level). In this way, PAM provides a systematic description of perceptual assimilation of non-native tones that includes both Categorised and Uncategorised responses¹, and allows for predictions about non-native perceptual assimilations based on both phonological and phonetic similarities to native phonemes.

According to PAM principles, native language phonology constrains perceptual assimilation of non-native phones to native categories. For example, if a non-native phonological contrast does not exist in the native language, naïve listeners may assimilate both members of the non-native contrast to a single native phonological category (Single-Category assimilation: SC) because they are not sufficiently sensitive to the non-native phonetic distinction. Strong native phonological effects on SC assimilation of non-native consonant and vowel contrasts have been reported in several previous studies. A

¹ We do not expect non-native tones to be assimilated by tone language listeners as nonspeech, i.e., as Non-Assimilable.

well-known example is that native Japanese speakers have difficulties categorising and 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). In addition, it was found that native American English (AE) speakers categorised two non-native bilabial stops that are phonologically contrastive in Zulu (voiced plosive /b/ vs. implosive /ɓ/) as nearly equivalent exemplars of the single native bilabial stop /b/ (Best, McRoberts, & Goodell, 2001).

Native versus non-native phonetic differences also influence non-native perception according to PAM principles. For example, although both AE and Japanese have a phonological contrast between /r/ and /w/, the phonetic realisation of the contrast differs in the two languages: /r/ is a central dorsal approximant [ɹ] in AE and an alveolar tap /ɾ/ in Japanese, and /w/ is lip-rounded [w] in AE but unrounded [u] in Japanese. Japanese listeners gave more /w/ responses than AE listeners in categorisation of an English /w/-/r/ continuum and their categorisation boundary was less steep because the Japanese listeners assimilated the rounded approximant AE /r/ as a poor exemplar of Japanese /w/ rather than as Japanese tapped /ɾ/ (Best & Strange, 1992). Similarly, when French listeners perceived an English /w/-/r/ continuum (/w/-/r/ is also contrastive in French), they categorised /r/ less consistently than AE listeners because /r/ in French is a uvular approximant or trill and thus differs phonetically from English /r/. Thus, they tended to report the AE /r/ as a /w/, like the Japanese did (Hallé et al., 1999). These and other analogous findings suggest that even when there is a similar *phonological* contrast in the listener's native language, *phonetic* differences from the native phonemes also affect perception of the non-native contrast.

1.2. Phonological features of Thai, Mandarin, Northern and Southern Vietnamese tones

The four tone languages used for the current study, the stimulus target language Thai (five tones) and three non-native listener languages, Mandarin (four tones), Northern Vietnamese (six tones) and Southern Vietnamese (five tones), differ phonologically in terms of the number and types of tones they contrast. Given that no universal cross-language phonological model of tone systems yet exists, we developed a parsimonious phonological feature system for our purposes of making phonological predictions for non-native tone assimilations. It includes tone contour and height features. The contour specifications are level (flat) contour and the dynamic contours rising, falling, and their combinations of falling-rising and rising-falling. The height specifications are high, mid, low. This system minimises dependence on specific, detailed phonetic realisations, and is capable of capturing tone distinctions for a wide range of tone languages (perhaps all). Contour (level or dynamic) is specified for each tone in a given language, but height is specified only when it is phonologically contrastive for tones of the same contour, e.g., height distinctions between two falling or two rising or two level tones in the same language. Note that among the four languages selected for the current study, the only height contrast for falling-rising contours occurs in Northern Vietnamese, and Mandarin tones (in cita-

tion form²) contrast only in contour type, with no minimal phonological height contrasts.

Thai, the target stimulus language, has three level tones contrasting in height (high, mid and low), and two dynamic contour tones that do not contrast in height (rising and falling) (Gandour, 1978). Mandarin has one level and three dynamic contour tones (a rising tone, a falling-rising tone, and a falling tone) in citation form (Chao, 1968).

The tone systems for the two dialects of Vietnamese are more complex. Northern Vietnamese has two phonologically level tones contrasting in height, high level *ngang* and low level *huyền*. It has four phonologically contour tones: rising *sắc*, falling *ng*, and falling-rising tones that contrast in height: high falling-rising *hỏi*, and low falling-rising *ngã* (Nhan, 1984). In the process of tonogenesis in Vietnamese (Haudricourt, 1954), in syllables that ended with a glottal stop, that stop evolved into the rising *sắc* and falling *ng* tones, which differ according to whether the syllable's onset consonant was voiceless versus voiced. Those two simple dynamic tones are phonologically distinct from the dynamically more complex falling-rising tones *hỏi* (high) and *ngã* (low) (determined by the voiceless versus voiced initials), which instead evolved from syllables ending with -s or -h. The reason for specifying *hỏi* as high and *ngã* as low is phonologically motivated (Yip, 2002): in tone reduplication in Northern Vietnamese, if the inputs are rising *sắc* or high falling-rising *hỏi*, the prefixal reduplicant surfaces as high level *ngang*, whereas if the inputs are falling *ng* or low falling-rising *ngã*, then the prefixal reduplicant surfaces as low level *huyền* (Yip, 2002).

Southern Vietnamese has five tones in its phonological system. Four of them correspond phonologically to Northern Vietnamese: high level *ngang*, low level *huyền*, rising *sắc*, falling *ng*. The fifth tone is reported to reflect a diachronic tone merger of the two falling-rising tones, high *hỏi* and low *ngã* (Brunelle, 2009), implying that the height distinction for falling-rising tones has been lost over time in Southern Vietnamese, which thus retains a height contrast only for level tones.

1.3. Phonetic characterisations of Thai, Mandarin, Northern and Southern Vietnamese tones

Several conventions exist for transcribing the phonetic details of lexical tones, including the diacritic convention used by the IPA (The International Phonetic Alphabet, 2015) and Chao numbers (Chao, 1930). As Chao numbers offer more flexibility in characterising dynamic contour tones and have been more widely used in studies of Asian tone languages, we relied on existing Chao transcriptions to provide a first approximation of phonetic characteristics of the tones in our selected languages. In Chao tone transcriptions the number 5 represents the highest pitch in tone production and the number 1 the lowest, and the sequence of Chao numbers indicates the pitch contour of the tone (Chao, 1930). We designated T for

² The focus of this paper is on perception of monosyllabic tones in citation form. We did not present tones in connected/conversational speech, as this is not typically done in perceptual assimilation studies, and moreover it introduces tone coarticulation and tone sandhi that affect the shapes of the tones. The latter effects are certainly of interest for future research but are beyond the scope of the current study, which already has a complex multifactorial design.

Thai, M for Mandarin, NV for Northern Vietnamese and SV for Southern Vietnamese. Thus, Mandarin high level tone is designated in Chao numbers as M55, rising tone as M35, falling-rising tone as M214 and falling tone as M51 (Chao, 1968).

The three phonologically level tones in Thai are not, however, all transcribed as being *phonetically* level according to their designated Chao numbers. The three Thai phonologically level tones are phonetically transcribed in Chao numbers as T45 (high level), T33 (mid level), T21 (low level) (Reid et al., 2015). The phonologically rising tone is phonetically transcribed in Chao numbers as T315 and the phonologically falling tone as T241 (Reid et al., 2015).

Northern Vietnamese phonologically level tones are phonetically transcribed in Chao numbers as NV44 high level (*ngang*) and NV22 low level (*huy n*), and the four dynamic contour tones as NV35 rising (*s c*), NV21 falling (*n ng*), NV214 high falling-rising (*hỏi*), and NV415 low falling-rising (*ngã*) (Nhan, 1984)³. Four of the five Southern Vietnamese tones are transcribed like their Northern Vietnamese counterparts: SV44 high-level (*ngang*), SV22 low-level (*huy n*), SV35 rising (*s c*), SV21 falling (*n ng*). The merged falling-rising tone (merger of NV *hỏi* and *ngã*) is transcribed as SV214 (Brunelle, 2009).

Note the non-level Chao phonetic transcriptions for two of the Thai phonologically level tones and the falling-rising and rising-falling Chao transcriptions for the phonologically rising and falling Thai tones, respectively. These cases highlight that there are discrepancies between the phonological characterisation of tones and the Chao transcriptions of their phonetic realisations. In addition, even the Chao transcriptions may also mismatch acoustic measures of the tones' actual contours. These observations highlight the difficulties of comparing the phonetic characteristics of tones across languages using Chao transcriptions alone. Therefore, we also conducted acoustic analyses of recorded tokens in each language to determine their detailed acoustic properties.

1.4. Acoustic properties of lexical tones

Lexical tones are realised primarily by pitch variations⁴ and are usually characterised acoustically as either pitch contours or discrete features derived from pitch contours. We followed that standard approach. We recorded production of two consonant-vowel syllables (/ma/ and /mi/) by four female native speakers of each language (Mandarin, $M_{age} = 27.0$ years; Northern Vietnamese, $M_{age} = 22.5$ years; and Southern Vietnamese, $M_{age} = 20.5$ years; Thai, $M_{age} = 30.3$ years) with each of their native tones. All target syllables are meaningful (free or bound) morphemes in the respective languages. There were 64 tokens (2 syllables /ma/ and /mi/ $\times$ 4 tones $\times$ 8 repetitions) for each Mandarin informant and 96 tokens (2 syllables /ma/ and /mi/ $\times$ 6 tones $\times$ 8 repetitions) for North and South Vietnamese

informants. Mandarin and Vietnamese productions were recorded using a Zoom H4n digital speech recorder with a sampling rate of 44.1 kHz 16-bit stereo format at a quiet testing booth at The MARCS Institute for Brain Behaviour and Development, Western Sydney University.

Mandarin items were elicited via Pinyin; Vietnamese and Thai items were elicited via the orthography of their languages in random order. Southern Vietnamese speakers were asked to read all six Vietnamese tones but were specifically instructed to read them in their southern accent, which allowed us to confirm whether the two falling-rising tones in Southern Vietnamese are indeed acoustically merged by our participant cohort. This merger was observed in our acoustic analyses of their tone productions (see Appendix B for details of modelling). In Northern Vietnamese, the acoustic differences between these two tones are significant $F(5.11) = 18.66$, $p < 0.001$ but in Southern Vietnamese, the difference is not significant, $F(2.00) = 0.16$, $p = .86$.

The Thai syllables were recorded for a separate study (Burnham, Kuratate, McBride-Chang, & Mattock, 2009), and were used here with permission from the authors. These materials were recorded in citation form in a sound-treated booth at The MARCS Institute for Brain Behaviour and Development, Western Sydney University, using a Lavalier AKG C417 PP microphone at the sampling rate of 48 kHz and 16 bit resolution. Many repetitions were produced by the speakers, but only natural-sounding high-quality exemplars of each token were selected. Five tokens per type were selected from two participants, four from the third participant, and six from the fourth participant for the current acoustic analysis. Thus, there were 20 tokens for each of the ten Thai target items (/ma:/ and /mi:/ $\times$ 5 tones each) for a total of 200 Thai tokens.

All boundaries of the recorded syllables were automatically marked and manually checked. The Praat script *ProsodyPro* (Xu, 2013) was used to provide syllable durations and 10-equidistant points of f0 values (in Hz). Raw f0 (in Hz) was normalised using the Lobanov (1971) method, which is commonly used in studies of, e.g., vowel acoustics across multiple speakers, which requires normalisation of frequency ranges and values across speakers. Lobanov-normalised f0 values reflect variation from the f0 mean of the speaker⁵. The most stable part of the normalised tone (from 10% to 90% of the syllable) was used to calculate all f0-related measures. Fig. 1 shows the Lobanov- and time-normalised mean f0 contours of lexical tones in Thai, Mandarin, Northern and Southern Vietnamese. For more details on discrete measures, such as syllable duration and f0-related measures, see Appendix A (Table A1).

1.5. PAM predictions

The basic PAM principle (Best, 1995) is that naïve listeners will assimilate non-native phones into their native phonological categories but with differences in goodness of fit due to the magnitude of phonetic discrepancies from the native phonemes. A non-native phone can be a good to moderate to poor fit (phonetic aspect, which is gradient) to a native phonological category (phonological aspect, which is categorical). Following

³ The designation of high versus low features in this contrast is based on their phonological behaviour in reduplication as stated in the phonological description section, and there is a discrepancy between the phonological features and Chao numbers in this case.

⁴ Other acoustic properties, such as voice quality, have been claimed to be associated with some tones in some languages, such as Vietnamese. For example, Pham (2004, 2003) includes phonation type as part of the phonological specification for some northern Vietnamese tones. However, these other features are not proposed to be required for tone languages generally nor specified for most tone languages that have been described. As our aim is to present a universal model of tone features, we focus on pitch properties, which are central to descriptions of every known tone language, in the present paper.

⁵ Calculating semitones instead, as suggested by one reviewer, would not have yielded the cross-speaker f0 normalisation needed for making cross-language tone system comparisons.

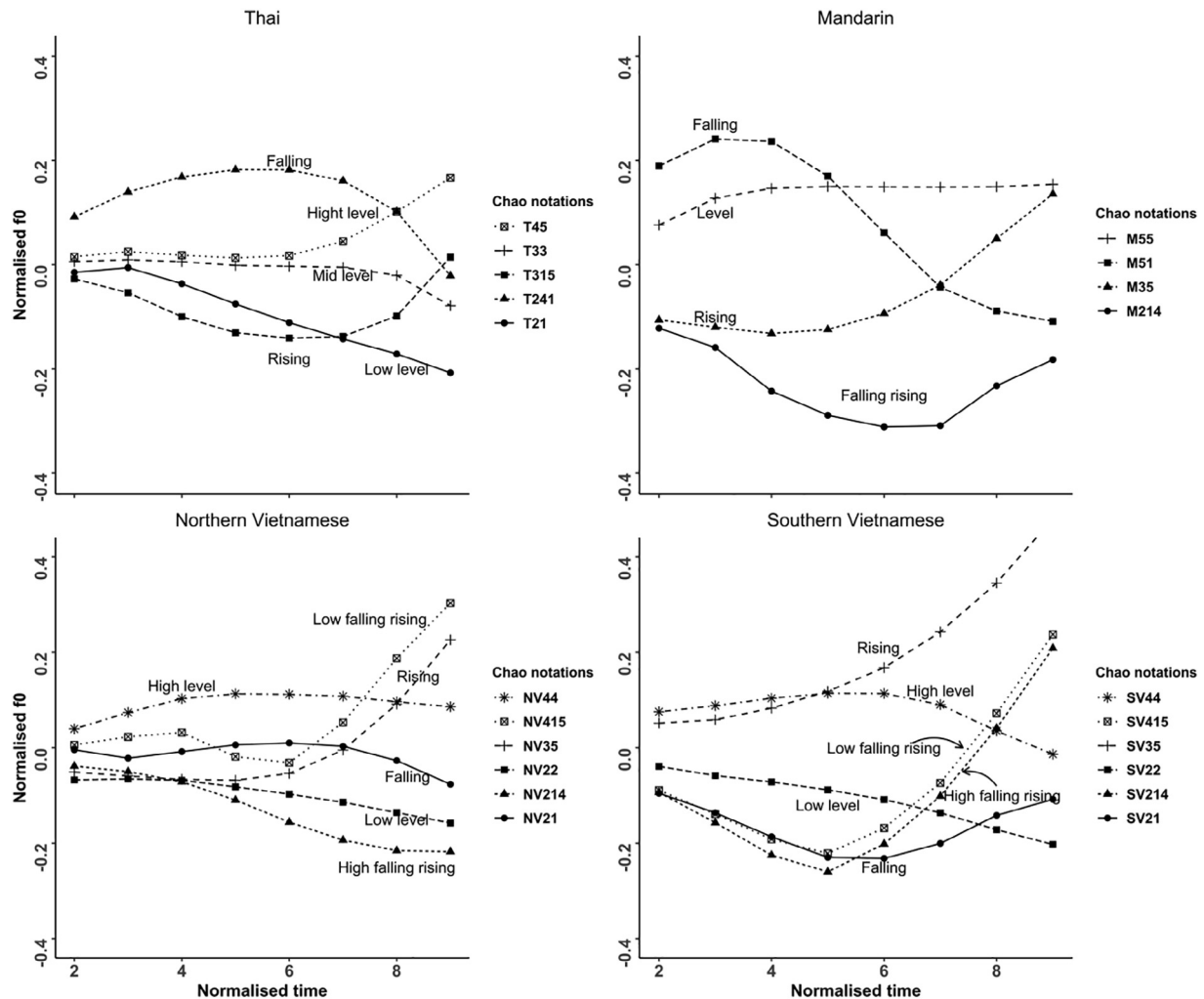


Fig. 1. Time-normalised and Lobanov-normalised mean f_0 contours of Thai (upper left), Mandarin (upper right), Northern Vietnamese (lower left) and Southern Vietnamese (lower right) tones, averaged across all tokens per tone type in each language. Thai: 20 tokens $\times$ 2 syllables (/ma:/ or /mi:/) $\times$ 5 tones; Mandarin: 32 tokens $\times$ 2 syllables (/ma/ or /mi/) $\times$ 4 tones; Northern Vietnamese/Southern Vietnamese: 32 tokens $\times$ 2 syllables (/ma/ or /mi/) $\times$ 6 tones. Phonological type labels are provided in the graphs for each contour.

these PAM principles, we posit that the extent of native phonological influence is reflected in the assimilation types. Phonological influence is generally strong for Categorised assimilations and is weaker for Uncategorised assimilations with variations for subtypes. In Uncategorised_{focalised} assimilations, the non-native phone is still assimilated as primarily similar to a single native category but choices of that native phoneme fall below the defined categorisation threshold and thus the phonological influence is moderate. In Uncategorised_{clustered} assimilation, the non-native phone is assimilated to a small set of native categories (below threshold but above chance level) and thus none of them have strong influence on assimilation. The general native phonological influence is therefore weak. In the Uncategorised_{dispersed} assimilation, a non-native phone category is spread across many L1 categories (all below chance level). The native phonological influence is very weak in this case.

Within those phonological constraints, listeners will nevertheless display some residual sensitivity to within-category *phonetic* variations from their native categories commensurate with the magnitude of phonetic discrepancy from the native tone(s) it is assimilated to. Residual native phonetic sensitivity is determined separately based on percent choice and good-

ness ratings of the chosen categories. We divided percent choice of the native tones above chance into three ranges: Low, Medium and High. These ranges respectively reflect strong, moderate, and weak residual phonetic effects. The percent choice ranges necessarily differ for the three groups. For Mandarin listeners, Low spanned 25–49% of choices, Medium 50–75%, and High 76–100%. For Northern Vietnamese listeners, Low covered 17–44%, Medium 45–73%, and High 74–100%. For Southern Vietnamese listeners, Low spanned 20–46%, Medium 47–74%, and High 75–100%. For the category-goodness ratings, we also divided the scale into three ranges, which apply to all listener groups: Low = 1–2.9, Medium = 3–4.9, and High = 5–7. These ranges reflect strong, moderate and weak residual phonetic effects, respectively.

High percent choice and/or high ratings reflect very strong perceived similarity to the corresponding native category and very weak residual phonetic sensitivity to within-category *phonetic* variations from their native categories. Low percent choice and/or low ratings, on the other hand, reflect very weak perceived similarity to the native tone and very strong residual phonetic effects. Percent choices reflect residual phonetic sensitivity in the categorisation process, which focuses on identifying how non-native phones may correspond to native

phonological categories. Goodness of fit ratings instead reflect residual phonetic sensitivity in the separate goodness rating process that focuses on evaluation of within-category phonetic differences between the non-native phone and native realisations of the category(s) to which it is assimilated. Thus, it is possible to observe discrepancy between percent choice and ratings within Categorised and Uncategorised assimilations, e.g., high-range percent choice with medium-range ratings or vice versa.

High and medium ranges of both percent choice and goodness ratings are predicted to occur for Categorised assimilations because they inherently entail perceived strong to moderately strong global phonetic similarities to the “most similar” native categories. Medium and low ranges of both percent choice and goodness ratings are predicted to occur for Uncategorised assimilations because the assimilated responses inherently presuppose moderately weak to much weaker perceived global phonetic similarities.

The only level of percent choice and ratings that overlaps between Categorised and Uncategorised_{focalised} (possibly also Uncategorised_{clustered}) is “Medium”, which would refer to Categorised cases that do meet the statistical criterion but with moderate levels of percent choice, as well as Uncategorised_{focalised/clustered} cases that fail the statistical criterion for Categorisation but still have moderate levels of percent choice for one or more native categories.

We extended these PAM principles to assimilation of non-native lexical tones by naïve tone language listeners to make predictions about non-native Thai tone assimilation by listeners of three tone languages/dialects. It is important to reiterate our assumption that naïve listeners lack phonological knowledge of the target language and thus are phonologically constrained only by their native tone system. Thus, we considered both phonological features and phonetic characteristics for the native tones but only the phonetic characteristics for the non-native tones.

T45 is phonetically rising in Thai and therefore should be Categorised as M35 because the general phonetic form of T45 fits abstractly the phonological feature specifications of M35 rising tone. The percent choice and category-goodness rating of M35 should be in the high range, reflecting a low residual sensitivity to non-native phonetic variations from the corresponding native category.

As noted earlier, Mandarin has only one level tone with no contrast in height. Thus, the Thai level tone T33 is expected to be Categorised as M55. The percent choice should be high which reflects low residual phonetic sensitivity in phonological categorisation to the only level tone in the native system, whereas the goodness rating should be medium as the phonetic evaluation of T33 in the rating task should reveal residual sensitivity to the moderate within-category phonetic deviation of T33 from M55 in terms of height.

T21 is phonetically low with a falling contour, but Mandarin does not have a phonologically low falling category. However, M214 is often realised phonetically as a low falling (21) allo-tone in non-final position although it acquires a final rise in the citation form (Yip, 2002). Thus, T21 should be Categorised as M214. However, the percent choice and rating for M214 as a response category will be medium given the notable departure from the native tone’s final rise in citation form (as in our

stimulus materials). M51 should also be selected as it is the actual falling tone in Mandarin. But the phonetic differences between M51 and T21 will lower the percent choice and category-goodness ratings to the low range.

T315 is phonetically falling-rising. Both M35 and M214 are phonetically similar to T315, in different ways, but M35 is more similar in terms of initial and final pitch height. Thus, it should be Categorised as M35 with medium percent choice and ratings, but M214 should be selected less often, with a low percent choice and category-goodness rating.

T241 is phonetically rising-falling and because Mandarin does not have such a phonological category, it is expected to be Uncategorised. Due to its greater phonetic similarities (height and contour) to M51 and somewhat to M55 (height) than to M35 or M214, as shown in Fig. 1, the percent choice of each of the former two categories should be higher than those for the latter two, resulting in Uncategorised_{clustered} assimilation. The percent choice and category-goodness rating of M51 should be medium, given the moderate phonetic similarity, and M55 choices and ratings should be in the low range.

For Northern Vietnamese listeners, T45 is phonetically rising and is similar to native rising tone NV35 but has a less steep rise with higher initial pitch and lower final pitch, so it should be categorised to NV35 but with medium percent choice and goodness ratings. NV415 is also similar to T45 in its initial and final pitch but NV415 has a dip in pitch in the middle of the tone whereas T45 lacks a medial dip in pitch. This phonetic difference from T45 should lead to some selection of NV415 but with low percent choice and low category-goodness ratings.

For Southern Vietnamese listeners, T45 should also be assimilated to SV35. However, SV214 (the merged falling-rising tone) and SV21 are also phonetically similar to T45 in terms of pitch contour, though with a lower overall pitch height (see Fig. 1). Given that SV214 and SV21 contours deviate from that of T45, and SV lacks a height contrast in contour tones, it is predicted that T45 will be Uncategorised and split among SV35, SV21 and SV214, thus Uncategorised_{clustered}/Uncategorised_{dispersed}. In this case, the percent choice of each of these SV tones should be in the low range, but the ratings should be in the medium range due to somewhat greater phonetic similarity.

Northern Vietnamese and Southern Vietnamese each have two level tones that contrast phonologically in height. T33 is phonetically lower than NV/SV44 but is only slightly above the height of NV/SV22, although it also has a somewhat different final contour. Thus, we predict the T33 would be Categorised as NV/SV22 with medium percent choice and category-goodness ratings. NV/SV44 should also be selected but the phonetic differences will lead to low percent choice and category-goodness ratings.

For low level T21, Northern Vietnamese and Southern Vietnamese both have a low level tone, NV/SV22, that is phonetically similar to T21. Thus, T21 should be Categorised as NV/SV22 with high percent choice and category-goodness ratings. The Vietnamese falling tones NV/SV21 are also phonetically somewhat similar to T21 but each of their contours differs phonetically from T21 as shown in Fig. 1, so they should be selected with a low percent choice and category-goodness rating.

T315 is phonetically falling rising and Northern Vietnamese has two phonologically falling rising tones, NV214 and NV415 but only NV415 has a final rise, as shown in Fig. 1. NV35 also has a similar final rise but without a dip in the middle of the tone. But NV415 is higher than NV35 overall and is steeper in the final rise of the contour and thus more deviant from T315 than NV35. Thus, T315 should be Categorised as NV35 with medium percent choice and goodness ratings, while NV415 will also be selected but with a low percent choice and category-goodness ratings.

On the other hand, Southern Vietnamese has only one falling rising tone, which reaches a low but not extremely low pitch in the middle. Thus, T315 is predicted to be Categorised as SV214 with medium percent choice and goodness ratings. SV35, with similar initial and final pitch but no dipping part in the middle of the contour, is less similar to T315 than SV214. It will be selected with low percent choice and low goodness ratings.

T241 is phonetically rising falling and both Northern Vietnamese and Southern Vietnamese lack rising falling tones in their tone inventories. However, NV/SV44 are phonetically similar to T241 in pitch height and throughout much of their contours (Fig. 1). Thus, T241 is predicted to be Categorised as NV/SV44; however, percent choice and category-goodness ratings should be medium.

Native phonological and phonetic influences will also be reflected in response times in the categorisation task. When a non-native tone is categorised to the native tone system, the incoming stimulus token is assumed to be stored in working memory (Baddeley & Hitch, 1974; Baddeley, 2010). If a non-native tone is phonologically Categorised and phonetically an ideal exemplar of the native tone, i.e., this tone falls squarely within that native phonological category, then the native category should be highly activated and receive a high percent of choices. No other native phonological categories have been activated and thus do not compete with it in the categorisation process. Thus, response times in this case should be short.

However, if a non-native tone is phonologically Categorised yet phonetically deviant from the native category, response times should be longer. That is, the processing cost should increase due to the phonetic discrepancy from the native tone, which would delay the decision (response time). In addition, if a non-native tone resembles two or more native categories, i.e., is assimilated as Uncategorised, then those multiple categories will be activated, each more weakly than in a Categorised assimilation. This should also impose an even greater processing cost due to the competition between the alternatives, which will delay the final categorisation decision. At the group level these added processing costs should yield selection of different (partially) activated native categories from trial to trial, as well as longer average response times due to increased category uncertainty.

Although the effects of talker variability on response time in speech processing have been examined (Antoniou & Wong, 2015), response time differences between Categorised and Uncategorised assimilation types have not previously been investigated. Within Uncategorised assimilation, as noted earlier, we further distinguish among three types of responses (*focalised*, *clustered* and *dispersed* responses), following Faris

and colleagues (2018). We predict that *focalised* responses, in which there is one major native response category, should yield relatively shorter decision times than *clustered* responses in which there are more than one native response categories activated and thus more competition. In a similar vein, *dispersed* responses should result in the longest decision times as most/all native categories are activated to a low and roughly equivalent level, that is, none of them are a good match for the non-native phone. The high phonological and phonetic uncertainty will maximise decision time and variability of categorisation choices.

2. Method

2.1. Participants

Twelve⁶ native speakers of Mandarin ($M_{\text{age}} = 29.6$ years, $SD = 6.1$; 10 females), Northern Vietnamese ($M_{\text{age}} = 20.4$ years, $SD = 2.5$; 6 females), Southern Vietnamese ($M_{\text{age}} = 23.7$ years, $SD = 8.6$; 10 females) participated in the experiment. Mandarin participants were tested at a university in Nanjing, China, except for three who were on a six-month academic visiting program in Sydney when they participated in the experiment. The Mandarin-speaking participants were all born and raised in various dialect regions in China (e.g., Fujian, Jiangsu, Jiangxi, Liaoning). All of their education had been conducted in Mandarin, spanning from early childhood through university, and they continued to use Mandarin on a daily basis. The Vietnamese participants were tested at universities in Sydney, Australia, where they were enrolled. Vietnamese participants were born and raised in the Northern regions (around Hanoi) or the Southern regions (around Ho Chi Minh city) of Vietnam and came to Australia mostly to study or migrate (Northern participants' $M_{\text{age of arrival}} = 19.7$ years, $SD = 2.3$; Southern participants' $M_{\text{age of arrival}} = 23.2$ years, $SD = 8.8$). According to a background questionnaire completed before the test, all had less than two years of formal musical training, which is important because more extensive musical training influences tone perception (Gottfried, Staby, & Ziemer, 2004). All participants self-reported to have normal hearing, no known speech-language difficulties, and none had experience with Thai⁷. The experiments were approved by the Western Sydney University Human Research Ethics Committee (HREC12560). All participants signed a consent form prior to testing and were compensated for their time (AU\$15 or the equivalent 60 RMB for participants in Nanjing, for 45 minutes).

2.2. Stimulus materials

We used a subset of the same monosyllabic Thai stimuli as in the acoustic analyses for the perceptual stimuli, with permission from the original authors (Burnham et al., 2009). Two syl-

⁶ We originally tested 13 participants in each group, but each group had one heritage speaker, so we conducted the analyses both with all speakers ($n = 13$ per group), and with the heritage speakers removed ($n = 12$ per group) for comparison. The only noteworthy difference is that the choice of M214 for T21 was lowered from 62.2% to 59.1% and that of M51 was raised from 25.7% to 27.9%, and consequently the difference between the two became marginal ($p = 0.06$). We reported results with a stricter control of listener backgrounds ($n = 12$ per group) but still considered T21 to be Categorised as M214.

⁷ They had begun learning English as school children in their home countries. Note that experience with English as a second language would not modulate their native tone system phonological knowledge, as English is not a tone language.

lables (/ma:/, /mi:/, with long Thai vowels) were chosen for target stimuli because they form (free/bound) morphemes for each native tone in Thai, Mandarin and Vietnamese, as noted earlier (1.4 *Acoustic properties of lexical tones*). We gave participants explicit instructions to assimilate the tones into their native tone categories (not to identify them as native words). This should minimise the effects of different morphological status. For the perceptual stimuli, we selected two tokens of each Thai target item by two native female speakers that had been judged by a third native Thai listener as the most natural sounding and correct.

2.3. Procedure

Participants were tested individually in a quiet room (e.g., sound-attenuated testing booth at Western Sydney University; library study booths at Macquarie University, University of Technology Sydney, and Nanjing University). Stimuli were presented on a Dell Latitude 7280 laptop running E-Prime Professional 2. Auditory stimuli were presented via Sennheiser HD 280 Pro headphones at 72 dB SPL.

Before the test session, participants completed 20 practice trials to familiarise them with the task and decisions they were asked to make on each trial. The stimuli in the practice trials were not used in the test session. No feedback was given in either the practice trials or test session.

The categorisation task had 120 trials (2 speakers $\times$ 5 tones $\times$ 2 tokens $\times$ 2 syllables $\times$ 3 repetitions). On each trial, the stimulus token was presented, and listeners made a forced-choice categorisation judgment to their native tones via a key press. Mandarin participants chose from four Pinyin options on stickers on keyboard keys, and Vietnamese speakers chose from six Vietnamese transcriptions on stickers on keyboard keys. Participants were asked to respond as quickly as possible within a 3 s response period. Responses beyond 3 s were not used for analysis; missing data due to time-outs account for only ~1% of the responses. Immediately following their categorisation decision, the same stimulus was played again in the same trial, and participants were asked to rate how well the tone fitted into the native category they had chosen, on a goodness rating scale (1 = Poor and 7 = Perfect). The ratings were used to index perceptual sensitivity to phonetic differences between the non-native token and the chosen native phonological category. Response times were also recorded for the categorisation responses.

3. Results

3.1. Categorisation criteria

Although in many published studies Categorised assimilation has been defined as the selection of one native category above an arbitrary pre-set threshold (Tyler et al., 2014) for a given non-native phone, we followed the tone categorisation criteria used in So and Best (2014) as being more appropriate for categorisation of lexical tones by listeners whose native languages differ in number of tones: first, a given native tone must be selected significantly more than chance level; second, that single native tone category must be chosen significantly more often than any other native categories. This method is

more sensitive to variations among different native tone systems than a rigid threshold is, as it considers the number of tone categories in a particular listener language.

For Uncategorised assimilations, we followed the subtypes established by the two previous studies of Faris and colleagues (2016, 2018), but modified their operational criteria to be based on our statistical approach to categorisation (Faris et al. had defined them using the fixed threshold approach) as follows: (1) for a *focalised* response, one non-native phone was considered as primarily similar to a single L1 category (above chance level) but choices of that native phoneme were not significantly higher than choices of other native categories; (2) for a *clustered* response, the uncategorised non-native phone was assimilated to two or more L1 categories above chance level, but they were not significantly different from each other; or (3) for a *dispersed* response, the choice of native phone category was made randomly across many L1 categories, all below chance level and not significantly different from each other.

Note that choices of SV214 (*hỏi*) and SV415 (*ngã*) (orthographically different) by the Southern Vietnamese group were combined for analyses of categorisations, ratings and response times, because they are phonologically merged rather than contrastive in Southern Vietnamese (see the evidence of merger in 1.4 *Acoustic properties of lexical tones*).

3.2. Percent choice and rating scores in perceptual assimilation

The mean percent choices and category-goodness ratings for each native tone response category are shown in Fig. 2 (for the full set of percent choices including those below 1% and the category-goodness ratings of below-chance choices, see Appendix C Table C1).

To determine whether an assimilation was Categorised or Uncategorised, we assessed whether categorisations of each Thai tone were significantly above chance level, which is 25% (1/4) for Mandarin, 16.7% (1/6) for Northern Vietnamese, and 20% (1/5) for Southern Vietnamese speakers, via a series of *t*-tests (see Appendix D, Table D1, for statistical details).

Then, for each listener group, mixed effects models were built for each of the five Thai target tones to determine whether the listeners assigned different native tone category labels to the Thai targets (details in Appendix D, Table D2). Participants were specified as a random factor. To calculate the *p*-values for the fixed effects, we used the Kenward-Roger approximation to the degrees of freedom, as recommend by Halekoh and Hojsgaard (2014), and the *Anova* function from the *car* package in R, with test specified as “F”.

Next, we ran multiple comparisons with the R-package *lsmeans* to determine which native category was selected more than other native categories that were above chance level (see Appendix D, Table D3 for statistical details). A Thai tone was considered Categorised only when one native category was selected significantly more often than all other native categories. We also ran *t*-tests on category goodness rating differences for cases where two non-native tones were assimilated into the same single native category. This allowed us to distinguish Single-Category assimilation (no significant difference in category goodness ratings) from Category-Goodness

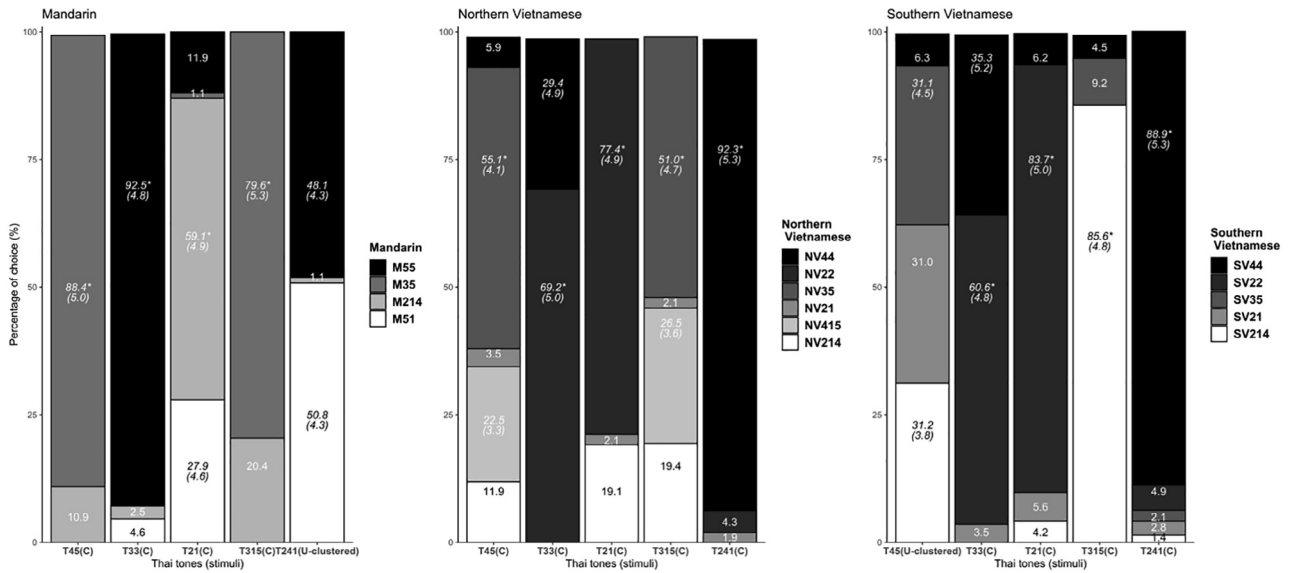


Fig. 2. Mean percent choices and goodness ratings (in the parenthesis) for each Thai tone by Mandarin, Northern Vietnamese and Southern Vietnamese listeners. Categories in italics are choices that were significantly above chance, which was 25% for Mandarin, 16.7% for Northern Vietnamese, 20% for Southern Vietnamese; “*” = Categorical tone. Assimilations: C = Categorical, U = Uncategorical. Ratings are shown for above chance level responses: 1 = poor, 7 = perfect. Native response categories less than 1% are not shown here but can be found in Appendix C, Table C1.

assimilation (significant rating difference between the two non-native tones), following PAM principles.

For Mandarin listeners, four of the five Thai tones were Categorical. Both T45 and T315 were Categorical as M35, but their category goodness ratings were not significantly different, thus meeting the Single-Category assimilation criterion. T33 was Categorical as M55. In all three of these cases only one native category was selected significantly above chance level and significantly more often than all other native choices. The fourth case, T21, was Categorical as M214, which was chosen marginally significantly more often than the other three Mandarin choices, although M51 was also chosen above chance. T241 was instead an Uncategorical_{clustered} assimilation because it was split evenly between M55 and M51, and both were chosen above chance level but not significantly different from each other.

For Northern Vietnamese listeners, all Thai tones were Categorical as native tones. T45 and T315 were both Categorical as NV35, and their category goodness ratings were not significantly different, meeting the Single-Category criterion. T33 and T21 were Categorical as NV22, and their category goodness ratings were not significantly different, again a Single-Category assimilation. T241 was Categorical as NV44.

Southern Vietnamese listeners Categorical both T21 and T33 as their native SV22, and their category goodness ratings were not significantly different, thus this pair forms a Single-Category assimilation. They also Categorical T315 as their SV214 and T241 as their SV44. However, T45 was an Uncategorical_{clustered} assimilation, as SV35 and SV214 were selected significantly above chance but not significantly different from each other.

3.3. Categorisation response times for the different assimilation types

The mean response times (ms) to each Thai tone for each native choice are presented in Appendix C Table C1. For formal statistical modelling, response time data were restricted

to the Categorical choices for the Categorical assimilations, and to above-chance choices for the Uncategorical_{clustered} assimilation. We predicted that the response time of Categorical assimilations should be shorter than that of Uncategorical assimilations. This is because as a non-native tone is assimilated across multiple native categories, these native categories will be activated, which should induce greater processing cost due to the comparison among the alternatives, and will delay the final categorisation decision.

We regrouped response time data across the three listener groups according to assimilation types, and ran a mixed-effects model with response time as the dependent variable and assimilation types (Categorical, Uncategorical) as the fixed factor. Participants were specified as random intercepts. A significant difference was found between Categorical and Uncategorical assimilations in terms of response time, $F(1, 146) = 14.96$, $p < 0.001$, suggesting that Categorical assimilations ($M = 764$ ms, $SD = 339$ ms) were significantly faster than Uncategorical assimilations ($M = 928$ ms, $SD = 348$ ms) as we predicted.

3.4. Residual native phonetic sensitivity

Next, we describe the range of percent choice and goodness ratings for the assimilations (see Table 1 for a summary in relation to our predictions).

First, we successfully predicted Mandarin listeners' Categorical assimilation of T45 as M35 and with accurate predictions of percent choice and rating ranges. T33, as predicted, was Categorical as M55 (the only level tone in Mandarin, not contrastive in height) by Mandarin listeners with a high percent choice, indicating a strong native phonological influence and reduced residual phonetic sensitivity in the categorisation process. However, in the rating task, Mandarin listeners did show sensitivity to phonetic differences from their native level tone as indicated by a medium level category-goodness rating.

In addition, as predicted, T21 was Categorical as M214 with medium range of percent choice and ratings. The assim-

Table 1
Summary of phonological and phonetic predictions and experiment results. Assimilations: C = Categorised, U = Uncategorised, H = High, M = Medium, L = Low, and (L) ranges for Categorised assimilations: for Mandarin speakers, L = 25–49%, M = 50–75%, H = 76–100%; for Northern Vietnamese listeners, L = 17–44%, M = 45–73%, H = 74–100%; and for Southern Vietnamese listeners, L = 20–46%, M = 47–74%, and H = 75–100%. For Uncategorised assimilation, the range is from Medium to Low and to Below chance. For ratings, Low = 1–2.9, Medium = 3–4.9, High = 5–7. Results higher than predictions are in boldface, indicating lower than expected within category residual phonetic sensitivity.

	Mandarin				Northern Vietnamese				Southern Vietnamese			
	Predictions		Results		Predictions		Results		Predictions		Results	
	Phonological	Phonetic (% rating)	Phonetic (% rating)	Results (% rating)	Phonological	Phonetic (% rating)	Phonetic (% rating)	Results (% rating)	Phonological	Phonetic (% rating)	Phonetic (% rating)	Results (% rating)
T45	C as M35	H	H	C as M35 (88.4 5.0)	C as NV35	M	M	C as NV35 (55.1 4.1)	UC _{clustered} /dispersed SV35 SV214 SV21	L	L	UC _{clustered} (31.1 4.5) (31.2 3.8) (31.0 4.7) C as SV22 (60.6 4.8) C as SV22 (83.7 5.0) C as SV214 (85.6 4.8) C as SV44 (88.9 5.3)
T33	C as M55	H	M	C as M55 (92.5 4.8) C as M214 (59.1 4.9)	C as NV22	M	M	C as NV22 (69.2 5.0) C as NV22 (77.4 4.9)	C as SV22	M	M	C as SV22 (60.6 4.8) C as SV22 (83.7 5.0)
T21	C as M214	M	M	C as M214 (59.1 4.9)	C as NV22	H	H	C as NV22 (77.4 4.9)	C as SV22	H	H	C as SV22 (60.6 4.8) C as SV22 (83.7 5.0)
T315	C as M35	M	M	C as M35 (79.6 5.3)	C as NV35	M	M	C as NV35 (51.0 4.7)	C as SV214	M	M	C as SV214 (85.6 4.8) C as SV44 (88.9 5.3)
T241	UC _{clustered} M51 M55	M L	M L	UC _{clustered} (50.8 4.3) (48.1 4.3)	C as NV44	M	M	C as NV44 (92.3 5.3)	C as SV44	M	M	C as SV44 (85.6 4.8) (88.9 5.3)

ilation of T315 to M35 was also successfully predicted but the percent choice and ratings were higher than expected, again indicating unexpectedly reduced residual sensitivity to differences from the native tone in both categorisation and rating processes, i.e., overridden by a strong native phonological influence.

T241 was predicted to be Uncategorised and split between M51 and M55. The M51 selections were expected to display medium range scores in percent choice and goodness ratings whereas M55 choices were expected to show low range scores in percent choice and goodness ratings. The Mandarin listeners in present study did split their responses to T241 between M51 and M55, yielding an Uncategorised_{clustered} assimilation, but percent choice and ratings were roughly equal between the two. This finding suggests that Mandarin listeners may have perceived the initial contour (24) in T241 as level rather than as rising. Previous reports have also shown some variations in assimilation of T241 by Mandarin listeners. Notably, previous observations encompass both of the assimilations we observed here: whereas both experienced (Thai learners) and inexperienced (Thai-naïve) Mandarin listeners in Wu et al. (2014) categorised T241 as M51, Mandarin participants in the study reported by Reid and colleagues (2015) categorised T241 as M55.

For the Northern Vietnamese listener group, our predictions based on PAM principles were also generally upheld. Northern Vietnamese listeners Categorised T45 as NV35, as predicted. The phonetic differences between T45 and NV35, as expected, lowered the percent choice and ratings to the medium range.

The prediction that both T33 and T21 would be Categorised to NV22 was supported. The percent choice of NV22 for T33 was smaller than that of T21, indicating larger residual sensitivity to T33 than T21 phonetic differences from NV22 in the categorisation process. The category-goodness ratings for both T33 and T21 were comparable, both reflecting relatively medium sensitivity to phonetic variations of the native categories from NV22 in the rating process.

We successfully predicted that T315 should be Categorised as NV35, with lower choices of NV415. Both the percent choice and category-goodness rating of NV35 were in the medium range, as we had predicted. NV415 was also selected as a native response category, but with low percentage of choice and medium category-goodness ratings. The rating was higher for NV35 than NV415, indicating that the residual sensitivity to phonetic differences from T315 was greater for NV415 than NV35.

T241 was Categorised as NV44 as predicted, but with higher than predicted percent choice and ratings. This reflects strong native phonological constraints and relatedly less phonetic sensitivity to the deviation of the non-native tone from the native one in both categorisation and rating processes.

Also consistent with our PAM-derived predictions, Southern Vietnamese listeners behaved quite differently from Northern Vietnamese listeners when assimilating the Thai rising contour tones T45 and T315. T45 was split by Southern Vietnamese listeners among SV35, SV214 and SV21, as expected. The percent choices were in the low range as predicted and the goodness ratings for all three native response categories were in the medium range, also as predicted. As we said, both SV21

and SV214 are phonetically falling-rising, and SV214 and SV415 have been subsumed into a tone merger (unlike the contrast maintained between Northern Vietnamese NV214—NV415: Brunelle, 2009). This difference between the two regional varieties of Vietnamese can account for the different assimilation patterns of T45 by Northern Vietnamese and Southern Vietnamese listeners.

On the other hand, as predicted, Southern Vietnamese participants Categorised T33 as SV22 in a similar way as Northern Vietnamese listeners but with lower percent choice and goodness ratings, suggesting stronger residual phonetic sensitivity to within category phonetic deviations of T33 from SV22 than Northern Vietnamese listeners. T21 was Categorised as SV22 by Southern Vietnamese listeners with similar percent choice and goodness ratings to those of Northern Vietnamese listeners.

T315 was Categorised into SV214 by Southern Vietnamese listeners with percent choice in the high range, higher than predicted (medium range), suggesting reduced residual sensitivity to phonetic differences between T315 and SV214 in the categorisation process, relative to the goodness rating process.

T241 was Categorised as SV44 but with higher percent choice and ratings than were predicted, suggesting reduced residual phonetic sensitivity to differences of T241 from SV44 in both categorisation and rating processes.

4. Discussion

The PAM-based predictions were generally upheld at both phonological and phonetic levels. Specifically, non-native listeners showed strong evidence of native phonological influences in categorisations, as predicted, as well as residual phonetic sensitivity to non-native tones that reflect native phonetic influences, that were taken into consideration in our extrapolation of PAM principles to perception of non-native tones by native listeners of other tone languages.

We argued earlier that native phonological factors are indicated by Categorised (strong native phonological influences) or Uncategorised assimilations (weak native phonological influences) to the listener's native tone system. Phonetic factors, on the other hand, are reflected in the relative percent choice in the categorisation process, and/or the goodness ratings in the rating process for a native tone category. We considered both the type of assimilation (Categorised versus Uncategorised) and the relative percent choice and goodness ratings to disentangle phonological versus phonetic influences from the native language on non-native tone categorisation.

In the experiment, we found only two Uncategorised assimilations whereas all other assimilations were Categorised which suggests that native language phonological constraints are strong in cross-language tone categorisation. Another piece of evidence for native phonological influences is that in some Categorised assimilation cases, the percent choice and/or category-goodness ratings were somewhat higher than predicted (seen in T33 → M55; T315 → M35 and SV214; T241 → NV/SV44). In these cases, it appears that strong phonological influences in Categorised assimilation also reduced residual sensitivity to phonetic deviations of the non-native tones from the native tones. This finding is consistent with the PAM principle that native phonological influences may hinder sensitivity to within-category phonetic differences.

It is also compatible with claims by NLM that tokens near a native prototype (most phonetically similar to it) are the most greatly attracted to the prototype with reduced sensitivity to phonetic differences, but the farther away the tokens are from the prototype (the more deviant they are from the prototype) the more their phonetic differences are detected.

Strong evidence for phonetic effects, i.e., strong residual sensitivity to phonetic variation from native categories, are indicated by low-range percent choice and/or goodness ratings (seen in T241 → M55, T45 → SV35/SV214/SV21). More interestingly, in some cases, we found a discrepancy between percent choice and goodness ratings. For example, Mandarin has only one level tone M55 and thus does not contrast level tone height phonologically as Thai does. Mandarin listeners consistently Categorised mid-level T33 into M55 with high percent choice, indicating reduced phonetic sensitivity in the categorisation process, but with medium goodness ratings, indicating moderate residual sensitivity to phonetic deviations of M55 from T33 in the rating process. This finding is also in line with the claim from SLM (Flege, 1995) that an equivalence classification of a non-native phone to a native phoneme category can override detection of second language (L2) features (pitch height in this case) that are not contrastive in the native language, but that listeners can still retain sensitivity to phonetic differences in some tasks.

Additionally, the differing assimilation patterns for T45 between the Northern and Southern Vietnamese groups reflect native phonetic influences. PAM made different predictions for T45 assimilation by the Northern (Categorised as NV35) and Southern Vietnamese groups (Uncategorised and split among SV35, SV214, SV21) as it considered the phonetic differences between Northern Vietnamese and Southern Vietnamese rising/falling-rising tones. As can be seen in Fig. 1, the phonetic trajectory and height of T45 is more similar to NV35 than to any other NV tones. On the other hand, T45 is roughly equivalently weak in similarity to the heights and trajectories of SV35, SV214, and SV21, the latter of which is realized with a slight final rise, unlike its Northern Vietnamese counterpart NV21.

Also as predicted, we observed longer response times for Uncategorised than for Categorised assimilations. We suggest that the Categorised cases indicate a strong, straightforward phonological influence from the native language tone system, where there is a single native category activated much more highly than other categories, with little to no phonological competition, resulting in fast responses. However, Uncategorised assimilations reflect weaker native phonological influence and involve different processes from the Categorised ones. The longer response time in Uncategorised assimilations likely reflects an extra processing cost and perceptual uncertainty, caused by competition among multiple weakly activated native categories and phonetic discrepancy from those native categories.

In summary, PAM accurately predicted assimilation types and/or percent choice and goodness rating ranges in the great majority of cases. We observed strong native phonological effects in Categorised assimilations, especially when percentage choice and goodness ratings were high. Strong phonetic effects reflected in residual sensitivity to phonetic variations within categories were indicated in low percent choice and/or goodness ratings; moderate phonetic effects were reflected in medium percent choice and/or ratings.

We acknowledge that PAM predictions typically focus on the relationships between assimilation patterns for non-native contrasts and discrimination levels for those contrasts. Discrimination was not addressed in the current study, which examined phonological and phonetic influences of different native tone systems on the listeners' assimilation patterns for unfamiliar Thai tones. We also wished to identify Thai tone contrasts for future research examining the PAM predictions for discrimination of tone contrasts based on the current tone categorisation findings. We examined the categorisation-discrimination relationship in a separate study (Chen, Best, Antoniou, & Kasisopa, 2019) in which we found that Mandarin listeners had greater difficulties in discriminating T45 and T315, which were Categorised into the same native category (Single Category assimilation), than in discriminating T21-T33, which were Categorised into different native categories (Two Category assimilation). It would be important to look at discrimination of the T21-T33 contrast by Southern Vietnamese listeners, who instead Categorised it into the same native category. According to PAM principles, they should discriminate that contrast significantly more poorly than the Mandarin listeners did.

Findings of assimilation patterns not only carry theoretical implications for non-native perception research, but can also shed light on second language speech learning (Best, 2019). For example, according to PAM-L2 (Best & Tyler, 2007), when a non-native (L2) phone is Categorised as a good exemplar of a L1 phonological category, no further perceptual learning is likely to happen. This can be beneficial to L2 learning when the native category is phonetically similar to the non-native category. However, Categorised assimilation could be counterproductive in L2 tone production accuracy if the non-native category is phonetically different from the corresponding native category. The current study did not examine L2 tone production and its relationship to perceptual assimilation of tones, which also addresses the core principles of SLM (e.g., Flege, 1995). Future research could do so by investigating non-native tone imitation. For example, assimilating T33 into M55 could lead Mandarin learners of Thai to produce T33 with an inappropriately high f_0 . However, if Mandarin listeners perceive T33 as functionally equivalent to M55 at the phonological level but perceive a phonetic difference between L1 and L2 tones, they may learn to refine the phonetic details of their production for the L2 tone. In a preliminary imitation study using the same stimuli (Chen, Best, & Antoniou, 2019), we found that Mandarin participants' imitation of T33 was accurate in terms of pitch height, suggesting that while they perceptually Categorised T33 as M55, the phonetic differences between T33 and M55 were retained in memory and available for imitation and L2 speech learning.

For non-native tones that are not Categorised as any single L1 phonological category but are heard as being similar to several L1 categories, i.e., Uncategorised assimilations, PAM-L2 predicts that one or two new L2 phonological categories may be formed. Similarly, SLM claims that new categories could be formed for this type of L2 phones, and that if the new phonetic category matches that of native speakers of the L2, then the L2 phone will be produced accurately. However, PAM-L2 predictions differ from SLM in that PAM considers the comparative relationships within the interlanguage phonological system in addition to the similarity of a given L2 phone to the closest individual native phonetic category. If the Uncate-

gorised L2 phones are assimilated into different sets of L1 phonemes with little overlap in the native categories chosen, PAM-L2 predicts that two or more new categories could be formed. But if the Uncategorised L2 phones are identified as similar to the same set of L1 phonemes, then only a single new category would be formed, and discrimination of the contrasting L2 phones may remain difficult (Best, Avesani, Tyler, & Vayra, 2019). Further research could examine the discrimination of non-native tone contrasts in which the two tones are assimilated to overlapping or non-overlapping set(s) of native tones by tone language listeners.

5. Conclusions

To conclude, our findings demonstrate that perceptual assimilation of non-native tones is affected by listeners' native phonological categories as well as the phonetic details of those categories. PAM predictions, which consider both phonological and phonetic factors in predicting non-native assimilation, were upheld. Naïve listeners Categorised many non-native tones into their native tone phonological categories while at the same time often retaining some degree of sensitivity to phonetic differences between the non-native and native tones. When there was only one corresponding native phonological category for a non-native tone, phonological constraints overrode phonetic differences in affecting percent choice in categorisation; however, listeners still retained residual phonetic sensitivity in goodness rating in some cases, indicated by lower than expected rating scores. Phonological and phonetic differences between two dialectal variants of Vietnamese affected assimilation types, reflecting native phonological effects, as well as percent choice and goodness-ratings, reflecting native phonetic effects. In addition, strong phonological influences facilitated Categorised assimilations, resulting in shorter categorisation response times than Uncategorised assimilations, which reflect weak native phonological influence and are more affected by non-native phonetic discrepancy and competition among alternative native categories.

The current findings have substantive implications for theories of tone perception as well as for second language lexical tone learning. Future research should consider the language-specific tone assimilation patterns found in the present study and compare discrimination performance among the same non-native tone contrasts that are assimilated differently by naïve listeners of different native language backgrounds. Furthermore, future research should also compare their imitations of non-native tones with their assimilations of those tones, to detect native language perceptual influences on production, as hypothesised by SLM and anticipated by PAM assumptions about perception of articulatory information in speech.

CRedit authorship contribution statement

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

Acknowledgements

This research was supported by a China Scholarship Council and Western Sydney University Joint scholarship awarded to the first author, Juqiang Chen.

Appendix A. Acoustic measures of tones in Thai, Mandarin, Northern Vietnamese and Southern Vietnamese

While normalised pitch contours offer qualitative visual comparisons of temporally dynamic characteristics of tones across languages, discrete features can allow quantitative comparisons, including duration (which was removed via time-normalisation for Fig. 1). Table A1 presents a summary of six acoustic measures for each tone per each language. According to a previous multidimensional scaling study, duration, and f0 mean, direction, extreme endpoint, and slope have been found to correlate with perception of tones by native listeners of Thai and Yoruba (Gandour, 1978). To capture these features, we calculated syllable duration, f0_{onset}, f0_{offset}, f0_{mean}, and f0_{excursion} (maximum to minimum), which had been used to characterise level tone contrasts in a previous study (Kuang, 2013), and we added one more measure, f0_{max_location} ratio, (i.e., relative location of the f0 peak as a proportion of the duration of the tone) to distinguish differently-timed peaks in convex and concave contours (e.g. T241 and T315).

Ideally it would be helpful to provide statistical comparisons based on these discrete measures by individual linear mixed-effect models, or principle component analysis or machine learning classification algorithms. We have conducted all of these on our acoustic data but each was accompanied by serious methodologically inherent problems as pointed out by reviewers. Significant differences in acoustic dimensions as shown in the linear mixed effects models and the Tukey adjusted multiple comparisons do not necessarily lead to problems in perceptual assimilations, which depend on listeners' relative weighting of the individual acoustic feature in their native language. Principle component analysis offers visual

comparisons, but its dimension reduction algorithm does not necessarily capture relative weightings in the same way as human listeners. Other machine learning classification algorithms, such as Support Vector Machine, could be applied but the interpretation would have to be treated very cautiously given our relatively small acoustic data set. Due to these issues and possible distractions of these models, we do not include the model-based comparisons we conducted, but we provide the measurements we took in the table below.

Appendix B. GAMM modeling for comparing Northern and Southern Vietnamese tone merger.

The tone merger of SV214 (*hỏi*) and SV415 (*ngã*) in Southern Vietnamese makes the phonological system of Southern Vietnamese different from that of Northern Vietnamese. Given the importance of this difference and its potential influences on perceptual assimilation patterns of Thai tones by these two groups, we verified the tone merger in Southern Vietnamese in terms of the pitch contours of these two tones in our acoustic study. In order to produce a formally holistic and dynamic comparison of f0 contours, we employed a General Additive Mixed Model (GAMM), which is a non-linear regression method that does not require aggregation or the pre-selection of a fixed time point in the contours. The method can detect general patterns over dynamically varying data, while at the same time accounting for subject and item-related variability (Nixon, van Rij, Mok, Baayen, & Chen, 2016; Wieling, 2018). In this way, it can uncover patterns that are obscured when data are aggregated or when a single time point is arbitrarily chosen.

Our dependent variable was the Lobanov-normalised f0 values which we modelled by using *smooths*. These smooths model non-linear patterns by combining a pre-specified number of basis functions. First, we built one model for each language/dialect separately. In each model, we set up smooths, *s(Time, by = tone)*, over time separately for each tone, *hỏi* and *ngã*. In order to take speaker variation into consideration,

Table A1

Acoustic measures of tones in Thai (20 tokens per tone), Mandarin, NV and SV (32 tokens per tone). All measures, f0_{mean}, f0_{onset}, f0_{offset}, f0_{excursion}, are Lobanov-normalised. * indicates that three decimal places were kept to show the real value was not equal to zero.

Tones	Duration (ms)		f0 _{mean}		f0 _{onset}		f0 _{offset}		f0 _{excursion}		f0 _{max_location} (%)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
T45	619	70	0.05	0.05	0.02	0.04	0.17	0.08	0.17	0.05	100	2
T33	640	85	-0.01	0.08	0.01	0.04	-0.08	0.09	0.11	0.07	34	23
T21	622	75	-0.1	0.04	-0.01	0.05	-0.21	0.1	0.22	0.1	22	7
T315	642	82	-0.08	0.05	-0.03	0.04	0.01	0.12	0.19	0.09	72	41
T241	565	101	0.13	0.06	0.09	0.05	-0.02	0.09	0.21	0.1	53	14
M55	663	137	0.14	0.04	0.08	0.07	0.15	0.05	0.09	0.05	74	27
M35	613	106	-0.05	0.05	-0.11	0.04	0.14	0.13	0.3	0.08	99	3
M214	745	137	-0.23	0.1	-0.12	0.08	-0.18	0.09	0.43	0.26	42	32
M51	506	127	0.08	0.06	0.19	0.14	-0.11	0.18	0.44	0.19	29	13
NV44	435	55	0.09	0.07	0.04	0.04	0.09	0.06	0.1	0.07	62	17
NV22	501	54	-0.1	0.08	-0.07	0.08	-0.16	0.09	0.11	0.06	25	12
NV35	431	61	*0.001	0.06	-0.05	0.04	0.23	0.14	0.31	0.13	99	3
NV21	286	50	-0.01	0.05	*0.004	0.04	-0.08	0.09	0.12	0.08	51	27
NV415	402	56	0.07	0.11	0.01	0.06	0.3	0.17	0.43	0.18	97	13
NV214	406	63	-0.13	0.06	-0.04	0.06	-0.22	0.07	0.21	0.07	18	9
SV44	469	37	0.08	0.04	0.08	0.03	-0.01	0.13	0.16	0.11	58	28
SV22	501	41	-0.11	0.05	-0.04	0.07	-0.2	0.07	0.17	0.06	16	8
SV35	465	41	0.19	0.05	0.05	0.05	0.48	0.12	0.44	0.1	100	2
SV21	456	40	-0.17	0.07	-0.1	0.1	-0.11	0.19	0.26	0.23	49	43
SV415	493	52	-0.07	0.06	-0.09	0.08	0.24	0.1	0.48	0.15	100	0
SV214	483	47	-0.1	0.06	-0.09	0.09	0.21	0.1	0.49	0.17	100	0

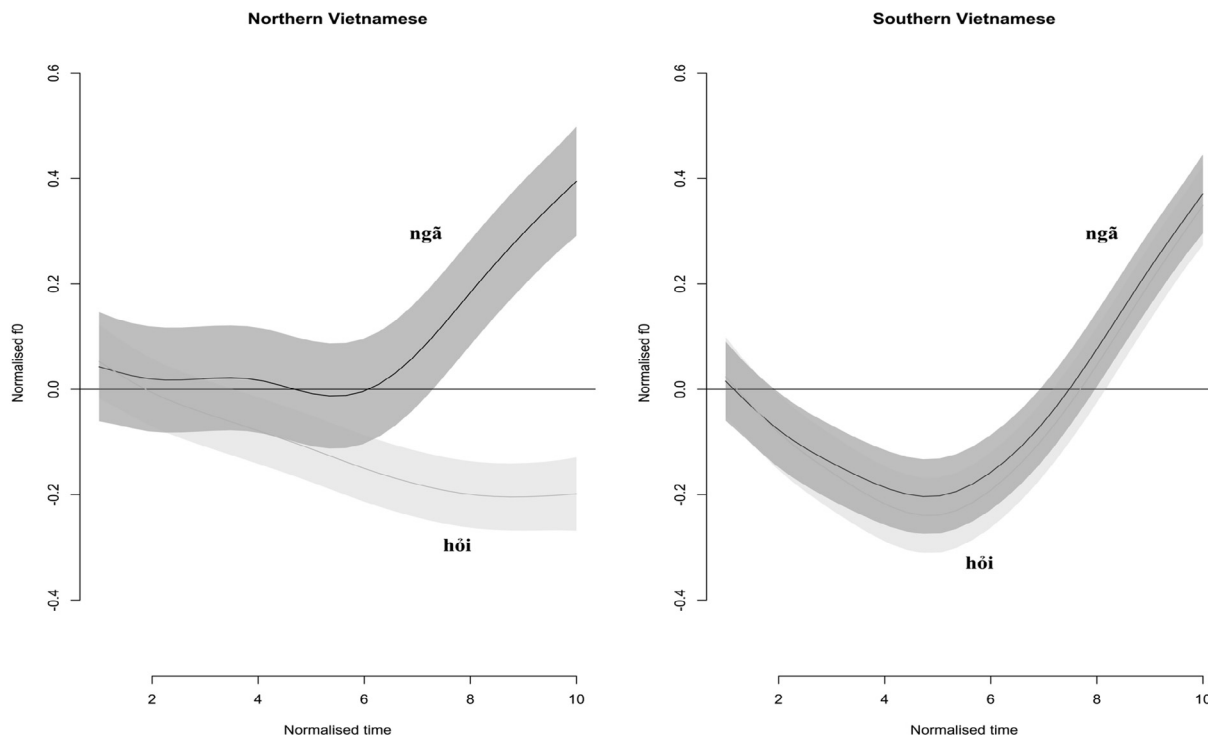


Fig. B1. Non-linear smooths for *hoi* and *ngã* in Northern Vietnamese (left) and Southern Vietnamese (right). The pointwise 95%-confidence intervals are shown by shaded bands.

we also modelled a non-linear random effect of speaker for each tone type via factor smoothing functions in the R package *mgcv*, as suggested by previous studies on articulatory data (Wieling et al., 2016; Wieling, 2018). $s(\text{Time}, \text{subject}, \text{by} = \text{tone}, \text{bs} = "fs", m = 1)$ is a factor smoothing function that models a non-linear difference over *Time* with respect to general time pattern for each *subject* (equal to random factors in the linear mixed-effect model). After we fitted the model, we used the *acf_resid* function in the *itsadug* package in R to obtain the autocorrelation measure and fed the autocorrelation measure at lag 1 to the *rho* setting in the new model. Fig. B1 shows the non-linear smooths of the tone contrasts in each Vietnamese regional dialect separately.

Visual comparison of the two panels in Fig. B1 suggests that the contrast exists in Northern Vietnamese but not in Southern Vietnamese. In order to formally compare the con-

trast in two Vietnamese dialects, we employed two approaches as suggested by (Wieling, 2018). The first approach is to refit the model with a binary difference smooth, which models the difference of the contrasts. In Northern Vietnamese, the difference is significant $F(5.11) = 18.66, p < 0.001$, but in Southern Vietnamese, the difference is not significant, $F(2.00) = 0.16, p = .86$. The second approach is to refit the model with an ordered factor difference smooth. The results also shows that the contrast is significantly different for Northern Vietnamese, $F(5.23) = 9.99, p < 0.001$, but not Southern Vietnamese, $F(1.01) = 0.10, p = .76$.

Appendix C

See Table C1

Table C1

Mean percentage of choice (%), mean category-goodness ratings and mean response times (RT, ms) for categorisations of each Thai tone to the tones in each listener language. "—" means no response. Ratings: 1 = poor, 7 = perfect.

Native categories	T45			T33			T21			T315			T241		
	%	rating	RT	%	rating	RT	%	rating	RT	%	rating	RT	%	rating	RT
M55	0.7	2	1869	92.5	4.8	800	11.9	4.2	1376	—	—	—	48.1	4.3	1090
M35	88.4	5	711	0.4	5	1244	1.1	4	1996	79.6	5.3	667	—	—	—
M214	10.9	4.4	1414	2.5	3.5	1250	59.1	4.9	947	20.4	4.4	877	1.1	4	1582
M51	—	—	—	4.6	3.5	1533	27.9	4.6	1053	—	—	—	50.8	4.3	860
NV44	5.9	2.6	1393	29.4	4.9	821	1	1.3	307	0.7	3.5	334	92.3	5.3	620
NV22	1	2	1628	69.2	5	704	77.4	4.9	799	0.3	3	982	4.3	5	1215
NV35	55.1	4.1	924	—	—	—	0.3	6	1127	51	4.7	800	0.4	1	1630
NV21	3.5	3.5	1290	1	2	1112	2.1	2.7	911	2.1	2.4	1263	1.9	1.6	2126
NV415	22.5	3.3	1440	0.3	1	1795	—	—	—	26.5	3.6	1207	0.4	3	2449
NV214	11.9	4.2	1596	—	—	—	19.1	3.9	1089	19.4	3.5	1054	0.7	1	2178
SV44	6.3	3.2	1030	35.3	5.2	875	6.2	4	1042	4.5	2.5	1072	88.9	5.3	743
SV22	0.3	5	1758	60.6	4.8	835	83.7	5	641	—	—	—	4.9	4.8	1003
SV35	31.1	4.5	1086	—	—	—	0.3	7	604	9.2	4.6	669	2.1	4.5	928
SV21	31	4.7	1541	3.5	4.4	1175	5.6	3.4	1102	0.7	4.5	499	2.8	4.2	717
SV214	31.2	3.8	1139	0.7	2	1617	4.2	5	553	85.6	4.8	1013	1.4	4.5	718

Appendix D. Statistical tests for determining Categorised and Uncategorised assimilations.

See Tables D1–D3

Table D1

Testing native category choices against chance level (significant results are in bold, $p < 0.05$).

Thai targets	Native tone choices	<i>t</i>	<i>df</i>	<i>p</i>
T21	M214	4.13	9	0.001
T21	M51	2.34	5	0.033
T241	M51	2.64	11	0.011
T241	M55	2.88	10	0.008
T315	M35	10.19	11	<0.001
T33	M55	14.66	11	<0.001
T45	M35	15.1	11	<0.001
T21	NV214	0.49	11	0.317
T21	NV22	11.73	11	<0.001
T241	NV44	20.26	11	<0.001
T315	NV214	1.06	10	0.157
T315	NV35	4.85	11	<0.001
T315	NV415	2.61	9	0.014
T33	NV22	11.54	11	<0.001
T33	NV44	2.63	11	0.012
T45	NV35	4.81	11	<0.001
T45	NV415	3.34	7	0.006
T21	SV22	8.06	11	<0.001
T241	SV44	11.38	11	<0.001
T315	SV214	7.78	11	<0.001
T33	SV22	8.87	11	<0.001
T33	SV44	4.04	10	0.001
T45	SV21	1.35	9	0.105
T45	SV214	2.25	8	0.027
T45	SV35	1.8	10	0.05

Table D2

Mixed effect models of native category choices for each Thai tone stimulus (significant results are in bold, $p < 0.05$).

Native language	Thai stimulus	<i>F</i>	<i>df</i>	<i>p</i>
Mandarin	T315	98.73	3	<0.001
Mandarin	T241	16.78	3	<0.001
Mandarin	T21	9	3	<0.001
Mandarin	T33	259.64	3	<0.001
Mandarin	T45	209.54	3	<0.001
Northern Vietnamese	T315	22.92	5	<0.001
Northern Vietnamese	T241	391.35	5	<0.001
Northern Vietnamese	T21	106.4	5	<0.001
Northern Vietnamese	T33	107.08	5	<0.001
Northern Vietnamese	T45	16.33	5	<0.001
Southern Vietnamese	T315	47.97	4	<0.001
Southern Vietnamese	T241	140.69	4	<0.001
Southern Vietnamese	T21	60.5	4	<0.001
Southern Vietnamese	T33	68.65	4	<0.001
Southern Vietnamese	T45	4.21	4	0.006

Table D3

Multiple comparisons between native category choices for each Thai stimuli type with Tukey adjustment (significant results are in bold, $p < 0.05$). Only the most relevant comparisons are listed here.

Thai stimulus	Native choice pairs	<i>t</i>	<i>df</i>	<i>p</i>
T21	M214–M51	2.62	33	0.060
T315	M35–M214	11.05	33	<0.001
T45	M35–M214	18.65	33	<0.001
T33	M55–M51	22.23	33	<0.001
T241	M55–M51	−0.27	33	0.99

Table D3 (continued)

Thai stimulus	Native choice pairs	<i>t</i>	<i>df</i>	<i>p</i>
T21	NV214–NV22	−13.89	55	<0.001
T315	NV35–NV415	4.12	55	0.002
T45	NV35–NV415	4.58	55	<0.001
T33	NV44–NV22	−10.32	55	<0.001
T241	NV44–NV22	33.2	55	<0.001
T21	SV22–SV21	12.04	44	<0.001
T315	SV35–SV214	−10.16	44	<0.001
T45	SV35–SV214	−0.01	44	1
T45	SV214–SV21	0.02	44	1
T45	SV35–SV21	0.005	44	1
T33	SV44–SV22	−5.48	44	<0.001
T241	SV44–SV21	18.75	44	<0.001

References

- Antoniou, M., & Wong, P. C. M. (2015). Poor phonetic perceivers are affected by cognitive load when resolving talker variability. *The Journal of the Acoustical Society of America*, 138(2), 571–574. <https://doi.org/10.1121/1.4923362>.
- Baddeley, A. (2010). Working memory. *Current Biology*, 20(4), R136–R140. <https://doi.org/10.1016/j.cub.2009.12.014>.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1).
- Best, C. (1995). A direct realist view of cross-language speech perception. In Winifred Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.
- 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>.
- Best, C., Avesani, C., Tyler, M. D., & Vayra, M. (2019). PAM revisits the articulatory organ hypothesis: Italians' perception of English Anterior and Nuu-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: In Honor of Ocke-Schwen Bohn* (pp. 13–40).
- 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. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, 109(2), 775–794.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 13–34). John Benjamins Publishing Company. <https://doi.org/10.1075/llt.17.07bes>.
- Bohn, O.-S., & Best, C. (2012). Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners. *Journal of Phonetics*, 40(1), 109–128. <https://doi.org/10.1016/j.wocn.2011.08.002>.
- 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>.
- Brunelle, M., Nguyễn, D. D., & Nguyễn, K. H. (2010). A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica*, 67(3), 147–169. <https://doi.org/10.1159/000321053>.
- 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>.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press.
- Chao, Y. R. (1930). A system of tone-letters. *Le Maître Phonétique*, 45, 24–27.
- Chen, J., Best, C., & Antoniou, M. (2019). Cognitive factors in Thai-Naïve Mandarin Speakers' imitation of Thai Lexical tones. *Proc. Interspeech*, 2019, 2653–2657. <https://doi.org/10.21437/Interspeech.2019-1403>.
- Chen, J., Best, C., Antoniou, M., & Kasisopa, B. (2019). Cognitive factors in perception of Thai tones by naïve Mandarin listeners. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th international congress of phonetic sciences*, Melbourne, Australia 2019 (pp. 1684–1688). Australasian Speech Science and Technology Association Inc.
- Erickson, D. (1976). *A Physiological Analysis of the Tones of Thai*. [Ph.D., University of Connecticut]. <https://search.proquest.com/docview/302781298/citation/77296413DF0D40C9PQ/1>.
- Erickson, D. M., Liberman, M. Y., & Niimi, S. (1976). The geniohyoid and the role of the strap muscles in pitch control. *The Journal of the Acoustical Society of America*, 60 (S1), S63. <https://doi.org/10.1121/1.2003454>.
- Erickson, Donna, & Abramson, A. S. (2013). F0, EMG and Tonogenesis in Thai. *Journal of Nagoya Gakuin University (Language and Culture): Collected Papers in Honor of Prof. Katsumasa Shimizu*, 24–1.
- Escudero, P. (2005). Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization. LOT.

- 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>.
- Faris, M. M., Best, C., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America*, 139(1), EL1–EL5. <https://doi.org/10.1121/1.4939608>.
- Faris, M. M., Best, C., & Tyler, M. D. (2018). Discrimination of uncategorized 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. (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.
- Gandour, J. T. (1978). The perception of tone. In V. A. Fromkin, 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>.
- Halekoh, U., & Hojsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models—the R package pbrtest. *Journal of Statistical Software*, 59(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/jpho.1999.0097>.
- Haudricourt, A.-G. (1954). De l'origine des tons en vietnamien. *Journal Asiatique*, 142, 69–82.
- Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "Perceptual Magnet Effect". In I. W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). York Press.
- 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>.
- 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/S0142716400009796>.
- 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>.
- Nguyen, V. L., & Edmondson, J. A. (1998). Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies*, 28, 1–18.
- Nhan, N. T. (1984). *The syllabeme and patterns of word formation in vietnamese* [PhD Dissertation]. New York University.
- Nixon, J. S., van Rijn, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: Eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language*, 90, 103–125. <https://doi.org/10.1016/j.jml.2016.03.005>.
- Pham, A. H. (2004). Vietnamese tone: A new analysis.
- Pham, A. H. (2003). The key phonetic properties of Vietnamese tone: A reassessment. 1703–1706.
- Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, 52(1), 37–52. <https://doi.org/10.3758/BF03206758>.
- Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *The Journal of the Acoustical Society of America*, 97(2), 1286–1296.
- 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>.
- Sagart, L., Hallé, P. A., de Boysson-Bardies, B., & Arabia-Guidet, C. (1986). Tone production in modern standard Chinese: An electromyographic investigation. *Cahiers de Linguistique - Asie Orientale*, 15(2), 205–221. <https://doi.org/10.3406/clao.1986.1204>.
- 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. <https://doi.org/10.1017/S0142716400001417>.
- Scopus.
- 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., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America*, 109(4), 1691–1704.
- The International Phonetic Alphabet. (2015). IPA Chart, <http://www.internationalphoneticassociation.org/content/ipa-chart>, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2015 International Phonetic Association.
- 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>.
- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, 70, 86–116. <https://doi.org/10.1016/j.wocn.2018.03.002>.
- Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., ... Baayen, R. H. (2016). Investigating dialectal differences using articulatory data. *Journal of Phonetics*, 59, 122–143. <https://doi.org/10.1016/j.wocn.2016.09.004>.
- 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>.
- Xu, Y. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis.
- 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>.
- Yip, M. (2002). *Tone*. Cambridge University Press.