Journal of Phonetics 93 (2022) 101151

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

The impact of phonotactic features on novel tone discrimination

Jonathan Wright *, Melissa Baese-Berk

University of Oregon, Department of Linguistics, 1290 University of Oregon, Eugene, OR 97403-1290, USA

ARTICLE INFO

Article history:
Received 10 December 2020
Received in revised form 5 November 2021
Accepted 26 April 2022

Keywords:
Novel tone perception
Non-native lexical tones
Phonotactic structure
Thai
Mandarin

ABSTRACT

Many studies have examined novel tone perception, but few have investigated the interaction between novel tone perception ability and phonotactic structure. We examined the discrimination of Thai low and mid tones by native Mandarin and native English participants across four syllable types (CCVV, CVV, VV, hums) to test the interaction between first language suprasegmental experience and phonotactic complexity on novel tone discrimination. We also tested the impact of unfamiliar consonant clusters and unfamiliar segments in the onset. Across syllable types, native Mandarin participants discriminated tones better than native English speakers. Further, discrimination ability was not impacted by phonotactic complexity. However, unfamiliar syllable structure impacted discrimination ability for native Mandarin participants in an unintuitive way. They discriminated tones significantly better in CCVV syllables. Unfamiliar segments in the onset, however, had a negative impact on tone discrimination. The presence of /ŋ/ onsets, which are illegal in English and allophonically permitted in Mandarin, significantly reduced tone discrimination accuracy. For native English participants, /ŋ/ onsets resulted in no discrimination between tones. These results suggest that the phonotactic structure of carrier words for tones interacts with L1 phonotactic experience in modulating novel tone perception ability.

1. Introduction

Many of the world's languages have lexical tone, where contrasting pitch patterns on lexical items carries semantic meaning (Maddieson, 2013; Yip, 2002). When an adult attempts to acquire an additional language that has lexical tones, their ability to perceive the tone categories varies as a function of their experience with suprasegmental features in other languages they know. The ability to perceive novel tone contrasts has been a topic of inquiry for sixty years. The primary debate has centered on the extent to which previous experience with lexical tone benefits a person during novel tone perception. More recently, however, studies have begun examining other factors that interact with L1 suprasegmental experience during novel tone perception. We add to this growing literature by examining the interaction between L1 suprasegmental experience and phonotactic structure during novel tone discrimination.

2. Background

As a child learns their first language, they progressively develop phonemic categories specific to that language (Eimas et al., 1971; Kuhl, 1987; Werker, 1989; Kuhl et al., 1992). They learn to differentiate between phonemic categories based on features including voice onset time, phonation, F1, F2, F0 (pitch) height, and F0 contour, which vary as a function of the specific language. As the child learns their first language (L1), they learn to differentiate phonemic categories using only features present in their L1. Consequentially, as they gain experience with the L1, the ability to discriminate between contrasts not present in their L1 is reduced. A commonly used example is the difficulty Japanese speakers face when discriminating between the English "r" and "l" (Goto, 1971; MacKain et al., 1981, Sheldon & Strange, 1982). For Japanese speakers this difficulty arises as English differentiates between two phonemes, but in Japanese a single phoneme utilizes the same phonological space (Sheldon & Strange, 1982). When a person learns a second language, they often attempt to map the phonological space of the L2 onto the available acoustic spaces in their L1. Several speech perception models have been presented to account for this

* Corresponding author.
E-mail addresses: jwright8@uoregon.edu (J. Wright), mbaesebe@uoregon.edu (M. Baese-Berk).

https://doi.org/10.1016/j.wocn.2022.101151
0095-4470/

mapping. The Perceptual Assimilation Model (PAM) (e.g., Best, 1995; Best & Tyler, 2007), the Speech Learning Model (SLM) (e.g., Flege, 1995), and the Second Language Linguistic Perception Model (L2LP) (e.g., Escudero, 2005) all predict that L2 learners will try to map L2 phonemic categories to L1 categories that are closest to them in native phonological space. However, this mapping can occur in various ways. For example, two target L2 phonemes may be mapped onto a single L1 category, as in the Japanese example above. This can create difficulty in learning the two competing phonemes. It is much easier when two L2 phonemes map onto two different L1 categories. However, for some cases there simply is no L1 category for the L2 phoneme to map onto and the resulting acquisition of the phoneme can vary widely from poor to excellent (Best, 1995).

Some languages, such as Mandarin and Thai, have tone categories, where pitch height and/or pitch contour differentiate categories. L2 learners with tonal L1s are able to map similar L2 tone categories onto L1 tone categories (Reid et al., 2015; Chen et al., 2018; Chen et al., 2019, Chen et al., 2020). Chen et al. (2020) investigated the ability of native Mandarin speakers to match Thai tones to Mandarin tone categories. Thai tones that were similar to Mandarin categories were more consistently matched to Mandarin tone categories than dissimilar tones. Results from other studies suggest that experience with L1 tone benefits an L2 learner's ability to discriminate between two or more novel tones in the target language (i.e. novel tone discrimination; Lee et al., 1996; Wayland & Guion, 2004; Wang et al., 1999). On the other hand, L2 learners with non-tonal L1s (e.g. English) do not have L1 categories to map L2 tones onto. Research suggests that novel tone discrimination is difficult for native English speakers (Kiriloff, 1969; Bluhme & Burr, 1971; Shen, 1989; Sun, 1998; Wang et al., 1999; Wayland & Guion, 2004; Reid et al., 2015), but training studies have shown that they are capable of learning to discriminate between tone categories (Chen & Pederson, 2017; Chen et al., 2019), forming tone categories (Kiriloff, 1969; Wang et al., 1999; Guion & Pederson, 2007), and using tone categories to learn new lexical meanings (Wong & Perrachione, 2007). The contrast of L1 tonal experience on novel tone perception is investigated in the present study by testing the perception of novel Thai tones by native English speakers and native Mandarin speakers.

The effect of the participants' L1 on novel tone discrimination has been studied in depth. However, while results from some studies suggest that the L1 of the participant predicts perception accuracy (Lee et al., 1996; Wang, et al., 1999; Wayland & Guion, 2004; Wayland & Li, 2008; Qin & Mok, 2013), results from other studies suggest that the L1 effect is not consistent (Francis et al., 2008; So & Best, 2010). The latter studies point out that predicting novel tone discrimination ability is more complex than a simple dichotomy between tonal and non-tonal L1s. Results from these studies suggest that prosodic factors in the L1 (Francis et al., 2008), and phonetic features of the target tones affect novel tone perception (So & Best, 2010). The current study adds to this literature by examining phonotactic factors affecting the perception of novel tone categories. Specifically, we look at the interaction between L1 background and syllable structure in predicting discrimination accuracy.

Some previous work has demonstrated that tone learning can be improved by pairing specific tones with specific syllables (see, e.g., Chan & Leung, 2020; Wiener & Ito, 2015, 2016; Wiener, Ito, & Speer, 2018, 2021). In this work, learners are better able to acquire novel lexical tones when those tones are learned in specific syllable-tone co-occurrence patterns. This work suggests that listeners jointly process tone and syllable information.

However, nearly all tone discrimination and tone learning studies use, almost exclusively, tokens comprised of consonant–vowel (CV) syllables with L1-like segments. The use of these types of stimuli results in tone perception studies that specifically avoid using segments or phonotactic structures that are not native to the participants' first language or use pseudowords to avoid negative effects from non-native phonological patterns (e.g., Wong & Perrachione, 2007; Chandrasekaran et al., 2010). This is often a practical decision, as CV syllables are often easy to record and manipulate, and controlling for segmental content reduces the number of variables included within the stimulus set. However, sometimes this methodological decision belies a theoretical assumption that phonotactic complexity and/or novel segments may impact learning. For example, Liu et al. (2011) explicitly state that CV tokens are used based on the hypothesis that phonotactic complexity, especially involving novel segments, inhibits the learner's ability to attend to tone. They chose legal English syllable structures so that participants would not be distracted by unfamiliar syllables. However, the question remains whether participants are actually distracted by unfamiliar syllable structures or segments and to what extent. In the present study, we test unfamiliar syllable structures by including tokens with complex onsets (CCVV). Mandarin does not have complex onsets and so if unfamiliar syllable structures matter, then we would expect that native Mandarin participants would do worse on CCVVs than on CVVs.

We also test illegal phonotactic structures by including trials with CVV tokens having /ŋ/ in the onset. In Thai /ŋ/ onsets are legal, but English phonotactic structure does not permit /ŋ/ in the onset. Thus, if illegal syllable structure matters for novel tone discrimination, then we would expect native English participants to have worse discrimination accuracy on /ŋVV/ trials than on /nVV/ trials. Predicting results for native Mandarin speakers on CVV trials is less clear. In Mandarin /ŋ/ can occur in the onset as an allophonic variant when no other onset is present (Duanmu, 2007). So, if native Mandarin speakers do worse on /ŋVV/ trials, it would suggest that although /ŋ/ onsets are legal allophonically, they may be processed differently than phonemically legal onsets, resulting in disruption during suprasegmental discrimination.

Furthermore, due to experience with the L1, listeners may also be endogenously oriented to focus on different features in the stimuli. For example, tone perception studies show that native English listeners weigh pitch cues differently than Mandarin listeners (Guion & Pederson, 2007). Also, Chen and Pederson (2017) found that when attention is directed to segments, English learners of Southern Min do not improve in tone discrimination. That is, when listeners are asked to attend to consonant contrasts during training, they fail to acquire novel tone categories. However, when using the same stimuli, listen-

ers are able to learn tone categories when asked to attend to tone, instead of initial consonants. It may be that native English speakers, due to lack of experience with lexical pitch, are endogenously oriented to direct their attention to segmental structure during auditory perception, leading to difficulty during tone discrimination and tone category formation during L2 acquisition. Thus, we ask whether trials composed of tokens with lower phonotactic complexity would permit greater endogenous orientation to F0 information, resulting in greater discrimination accuracy. If so, we would expect to see an inverse linear relationship between phonotactic complexity and tone discrimination accuracy (i.e. discrimination scores on CCVV trials would be lower than trials composed of monophthong tokens (V)). The hypothesis that tone discrimination will be easier in less phonotactically complex syllables also has a foundation in the composition of the world's languages. There is a significant negative correlation between syllable structure complexity and the complexity of tone systems in the world's languages (Maddieson, 2007). This may suggest that in general, humans perceive tone more easily in the presence of lower phonotactic complexity.

The current study also contains naturally produced hummed tokens (hums). By including hummed tokens, we are able to investigate whether even further reduction in syllable complexity could modulate tone perception. The inclusion of these tokens is motivated by significant previous work that has investigated hummed tokens and tokens with segments. Van Lancker and Fromkin (1973) used hums to test Thai participants' discrimination of Thai tones in "non-linguistic" tokens and found that listeners processed hums differently from tokens with segments. They concluded that tones in hums are processed as non-speech sounds. Similarly, Jia et al. (2013) used dichotic listening to test Cantonese participants' discrimination of tones across tokens with segments and hums. They failed to replicate a significant difference in the cognitive processing of hums and tokens with segments. Overall, there have been relatively few studies on novel tone perception that use naturally produced hums. Therefore, while the suggestion that hums are perceived differently from tokens with segments continues to be a theme in speech perception studies, there is no overwhelming evidence from previous research that would lead us to conclude that the participants in the current study are perceiving hums differently from the other token types.

An issue that is even less clear is whether a difference in the processing of hums compared to the other token types would result in reduced or improved accuracy. Results from Van Lancker and Fromkin (1973) do not suggest that participants were more or less accurate on hums than on other token types. Jia et al. (2013), however, found that the discrimination of tones in hums was significantly worse than in tokens with segments. Overall, there is little evidence from research on the cognitive processing of novel tones in naturally produced hums to permit the development of a hypothesis and interpretation of the results in the current study. However, Jia et al. (2013) presented a hypothesis that is more measurable. They hypothesized that the lower accuracy on hums was due to the lack of acoustic complexity in hums compared to tokens with segments, stating that hums likely have fewer acoustic clues that could be used to discriminate tones. In our own stimuli, as shown in Section 3, hums do contain fewer acoustic cues that could be used to discriminate tones than tokens with segments. Thus, it is unclear whether the cognitive processing of hums differs from tokens with segments, and the current study does not speak to this question. However, it is clear that hums can contain fewer acoustic cues than tokens with segments. Therefore, by comparing the discrimination of tones in hums and tokens with segments, we test whether the reduction of acoustic cues that are normally present in tonal information reduces discrimination accuracy. Based on results from Jia et al. (2013), it is likely that tone discrimination in hums will be worse than CCVV, CVV, or VV syllables[1].

In the remainder of the manuscript we present an experiment designed to examine whether phonotactic structure interacts with tone perception for English and Mandarin listeners. In Section 3, we provide an acoustic description of our stimuli. In Section 4 and Section 5, we discuss the experimental methodology and results. In Section 6, we discuss the implications of these findings.

3. Acoustic description of the stimuli

The current study uses naturally produced tokens that include various syllable structures. CCVV and CVV syllables both use two different onsets. This variability in our choice of tokens, as well as inherent variability across productions of the tokens present acoustic variations that may impact discrimination ability, but may not be tied directly to F0. Therefore, it is beneficial to discuss differences in the acoustic signal that arise from the chosen segments and syllable structures and how those differences might impact the results from the study. We begin by describing the methodology used to create the stimuli.

Stimuli were recorded by two female native Thai speakers, age 26 and 31. Recordings took place in sound attenuated booths using a Shure SM35-XLR headworn microphone recorded onto a Zoom H4N Pro audio recorder. Talkers said the target tokens three times using the carrier phrase [kʰău phûːt kham wâː ___ sɔ̌ŋ khráŋ] 'they say the word ___ two times'. After transferring files to a computer, each token was extracted from the carrier phrase using Praat (Boersma & Weenik, 2015). Four recordings of each token type were made, and two recordings of each type were selected for each talker, choosing the two most similar recordings of the token type, with duration and similarity in pitch contour being primary factors. The chosen tokens were tested with five native Thai speakers, who were able to identify the tones used in each token with 88.3% accuracy (SD = 8.4). This accuracy score is high, considering that hums and nonwords were included and that the mid and low tones are the most challenging pair for Thai speakers to differentiate (Abramson, 1976). These

[1] It is important to note that while some work has suggested that hums are processed quite differently from linguistic stimuli (e.g., Van Lancker & Fromkin, 1973), other work has suggested that suprasegmental contrasts on hummed utterances can be linguistically meaningful (Gardner, 2001). Indeed, when listeners expect to process linguistically meaningful content, they can do so even with quite artificial and unnatural stimuli (e.g., Remez et al., 1981). Given that the hums in the present study are naturally produced and are presented along with more linguistically-typical stimuli (i.e., stimuli with consonants and vowels), we have no reason to believe, a priori, that these stimuli ought to be processed differently from the other stimuli in the present study.

stimuli provided two stimulus sets per talker for a total of 24 tokens per talker – 48 altogether.

The stimulus set consisted of CCVV syllables, /kʰraː/ and /plaː/, CVV syllables, /naː/ and /ŋaː/, VV syllable, /aː/, and a hummed syllable. To create minimal pairs in each of these types, standard Thai mid and low tones were used as they are acoustically similar and are known to be difficult to discriminate in isolation (Wayland & Guion, 2004; Wayland & Li, 2008). Fig. 1 illustrates the tone contours from the stimuli used in the experiment with normalized time. The contours represent means extracted from every ten percent of the duration of the tone bearing unit across all tokens. The initial and final portions of the durations were not used as they included large numbers of missing and randomly jittered values due to transitions to silence. F0 values were extracted using Praat (Boersma & Weenink, 2015), and the values were carefully inspected. Values were sometimes missing or randomly jittered due creaky voice, which occurred predominantly on lower F0 contours. These values were then measured manually. The light grey areas bordering the F0 contours represent ± 1 standard error of the mean.

The use of mid and low tones with each syllable type resulted in a necessary use of some nonce words to complete the stimulus set, specifically in the CCVV syllables, where low tone does not naturally occur. It is possible that the lack of lexical status of the CCVV + low tone syllables resulted in acoustic differences in the signal, particularly in the acoustic realization of the initial cluster. This may have contributed to the variability present in the CCVV tokens, and the variability may have affected tone discrimination accuracy on CCVV syllables compared to CVV and VV syllables, which had fewer issues regarding lexical status. Our analysis of CCVV syllables did not find acoustic differences that could be directly attributed to lexical status. However, when tokens were tested with native Thai speakers, CCVV syllables were identified with the highest accuracy (96.3%; CVV = 83.8%, VV = 87.5%, hums = 82.5%).

During the creation of the stimulus set, tokens were normalized for peak intensity. The durations of tokens were also equalized to within 0.15 seconds within trial by shortening longer tokens. Acoustic characteristics of the stimuli, including F1, F2, and F3 average and midpoint, and intensity and F0 mean, min, max, and range, were inspected for potential differences that would make any syllable type or individual token easier or harder to perceive. While there were no clear acoustic differences between the stimuli in these measurements, our investigation of the stimuli resulted in the identification of the following differences, which potentially could affect discrimination ability.

As mentioned, hums are used in tone perception studies as they reduce acoustic information to F0 without presenting additional segmental information. An acoustic investigation of the stimuli used in the current study reveals that hums did reduce acoustic cues compared with tokens with segments. All hums were produced without creaky voice, which is identified as irregular glottal pulses identifiable in the waveform or spectrogram. By contrast, in tokens with segments, creaky voice sometimes occurs concurrently with low tones. This distinction is shown in Figs. 2 and 3, where creaky voice is only present in /aː21/. F0, which is illustrated by the dotted line, is disrupted in /aː21/ by creaky voice towards the end of the utterance. Creaky voice is known to occur naturally on low tones (Davison, 1991; Thepboriruk, 2009) and can be used by native listeners to discriminate tones (Yu & Lam, 2014). However, the extent to which non-native listeners utilize creaky voice in novel tone discrimination is unclear. By including hummed tokens, we examine the impact that differences in the acoustic signal, particularly between phonation types, has on novel tone discrimination. One possible result is that listeners will do worse on hums compared to other syllable types due to the presence of creaky voice on low tones in CCVV, CVV, and VV tokens, which provides an additional acoustic cue to tone differences. Results from Cao (2012) suggest that native English listeners are impacted by creaky voice during tone identification much more than native Mandarin listeners. Therefore, it may be that native English listeners will exhibit a greater difference in discrimination ability between hums and other token types compared to native Mandarin listeners. However, results from Cao (2012) also indicated that neither group were able to use creaky voice to improve in tone identification accuracy. Therefore, it is also possible that the absence of creaky voice on hums would not result in lower discrimination scores for either group.

Potential differences in discrimination accuracy may also arise from variations in the acoustic signal due to the various onsets used. Duration, intensity, F0, F1, F2, and F3 across the stimulus set are discussed in Section 3.2 Stimuli. However, variations in CVV and CCVV syllables are worth describing in more detail as they may impact suprasegmental discrimination accuracy. For CCVV syllables /kʰraː/ and /plaː/ were used. Fig. 4 shows /plaː21/ and /plaː33/ with, as expected, voicing on /l/ and thus F0 information occurring after the initial voiceless stop. All /plaː/ tokens were identical in this aspect. However, /kʰraː/ tokens were highly variable.

In Thai the production of /r/ varies widely across speakers and within speaker and can be produced as [ɾ], [ɹ], [ɻ], [r], or [l]. Further, in everyday Thai speech /r/ is also regularly omitted, being realized as a singleton. However, in the tokens used in the current study /r/ was not realized as a singleton. Fig. 5 shows three of the /kʰraː/ tokens used in the current study. In the two /kʰraː21/ tokens the allophone [ɻ] is used, which is apparent in the spectrogram by the duration of the segment, the lowered f3, and the lack of F0 information. Fig. 5 illustrates variability across /kʰraː/ tokens used in the current study. For example, the duration of [ɻ] differs between the two /kʰraː21/ tokens. In the /kʰraː33/ token in Fig. 5, the allophone [r] is used, which provides some F0 information. Other /kʰraː/ tokens in the stimulus set use the allophone [ɾ]. Each allophone occurs with mid tones and low tones. Therefore, CCVV syllables contain the greatest amount of acoustic variability in the stimulus set used in the present study. As previously discussed, this type of variability in the acoustic signal is specifically controlled in most tone perception experiments as it is expected that such segmental variability would hinder suprasegmental perception. And so, we may expect in the current study that novel tone discrimination accuracy will be worse in CCVV syllables.

For CVV syllables /naː/ and /ŋaː/ were used. Fig. 6 shows /ŋaː21/ and /naː21/, and Fig. 7 shows /ŋaː33/ and /naː33/. Both /ŋaː21/ and /naː21/ were produced with creaky voice, while /ŋ

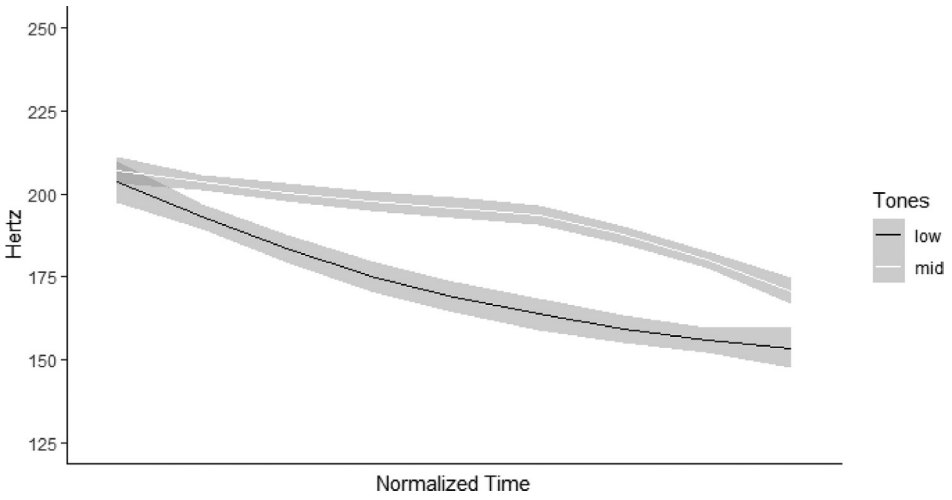

Fig. 1. Mean F0 contours and ± 1 standard error of the mean for mid and low tones across normalized time.

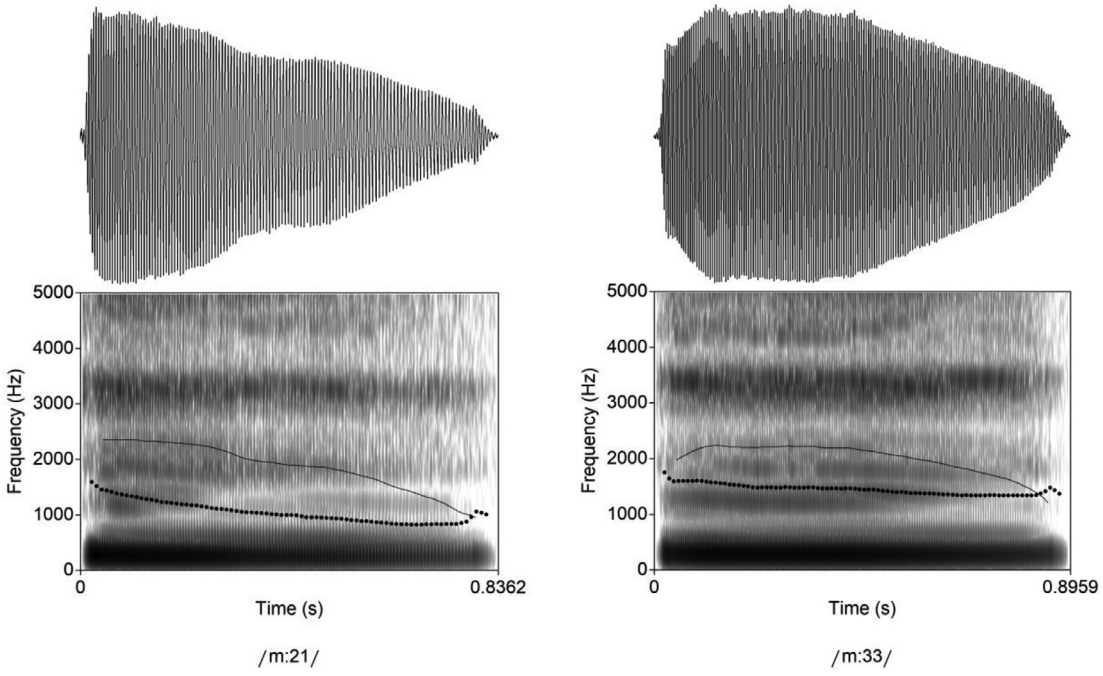

Fig. 2. Thai mid and low tones in a hummed token, showing no creaky phonation on mid or low tones.

aː33/ and /naː33/ were not. On all four stimuli F0 information occurs throughout the onset. Overall, there are no distinguishing acoustic features in the onset of /ŋaː/ and /naː/ syllables that might suggest one onset type would be easier to discriminate tone in than the other. Thus, from a bottom-up perspective, performance on /ŋaː/ and /naː/ syllables should be identical. Also, it may be expected that tone discrimination accuracy would benefit from F0 information being carried across the onset over other syllable types that have no onset (VV and hums) or have voicing disfluencies in the onset (CCVV). However, this possibility seems uncertain with the tones used in the current study as both tones begin at the same pitch height and only diverge after approximately 25% of the utterance. Thus, the presence of F0 information in the onset would not provide the listener with contrastive tonal information.

4. Methods

4.1. Participants

Twenty-nine native English speakers (8 male, 21 female) between 18 and 30 years old and thirteen native Mandarin speakers (7 male, 6 female) between 18 and 31 years old participated in this experiment[2]. Participants were undergraduates recruited from the University of Oregon Psychology and

[2] Given a variety of restrictions including limitations related to COVID-19, we were unable to control for language background of our Mandarin speakers. All of them reported being native Mandarin speakers and all are proficient in English, as is evidenced by their admission to UO, which did not require additional English language courses. We do not have reason to believe that the language backgrounds of our Mandarin participants systematically impacted their performance, but future work could investigate the role of specific home dialect of Mandarin on perception of Thai tones.

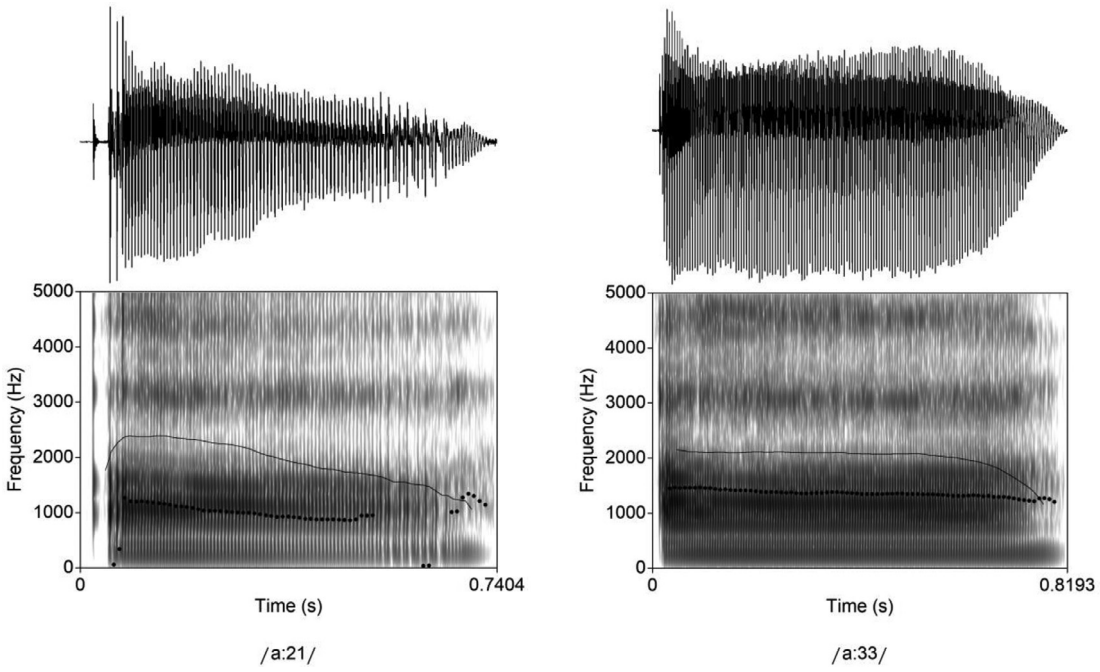

Fig. 3. Thai mid and low tones in the token /a:/, showing creaky phonation on the low tone only.

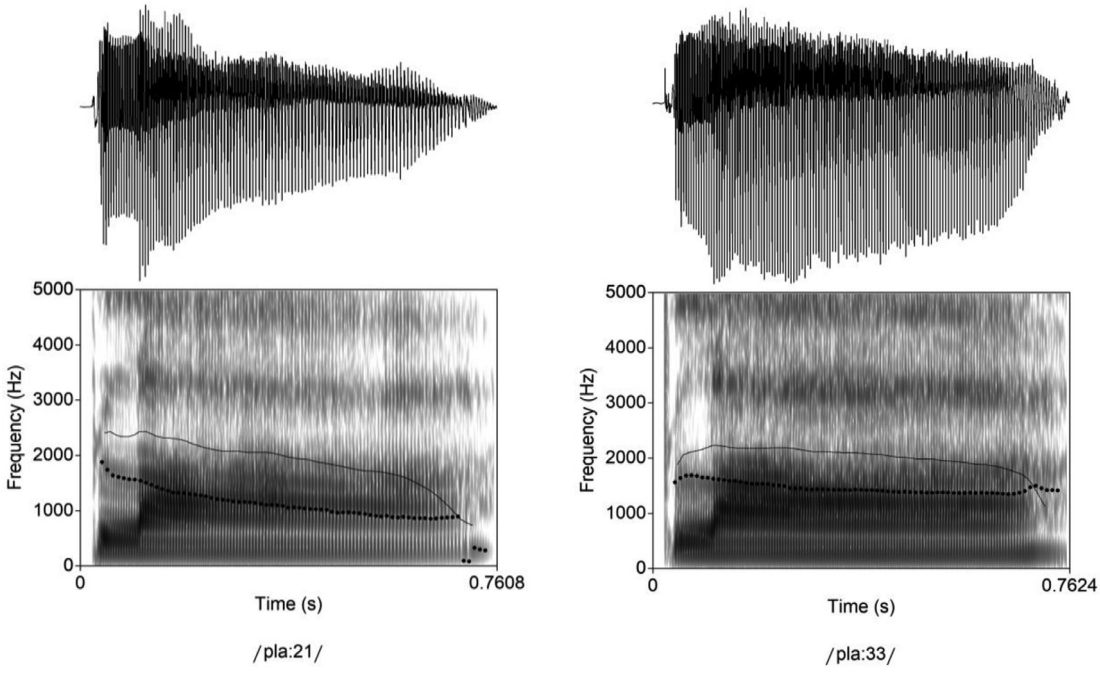

Fig. 4. CCVV tokens /pla:21/ and /pla:33/, illustrating invariance in the initial cluster.

Linguistics subject pool and received partial course credit for their participation. Three native English speakers with severe hearing impairments were removed from the study.

4.2. Stimuli

The stimulus set is discussed in Section 3. Twenty-four unique tokens per talker were used, for a total of 48 unique tokens.

4.3. Procedure

An AX ("same-different") test was used to determine tone discrimination accuracy across syllable types for the two groups. The two tokens in each trial either had the same tone or different tones. To increase the difficulty of the task, each token within a trial was produced by a different talker. Wayland and Guion (2004) and Chen et al. (2019) used different speakers within trial to increase difficulty so that listeners would not demonstrate ceiling effects. Chen et al. (2019) also

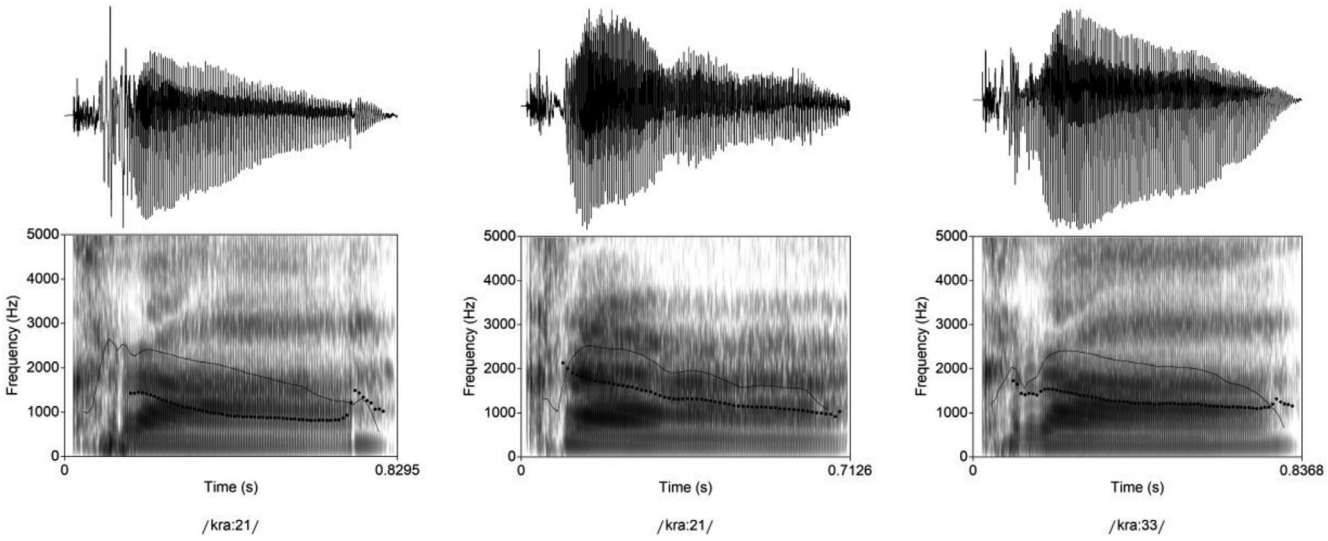

Fig. 5. CCVV tokens /kʰra:21/, /kʰra:21/, and /kʰra:33/, illustrating variability in the initial cluster.

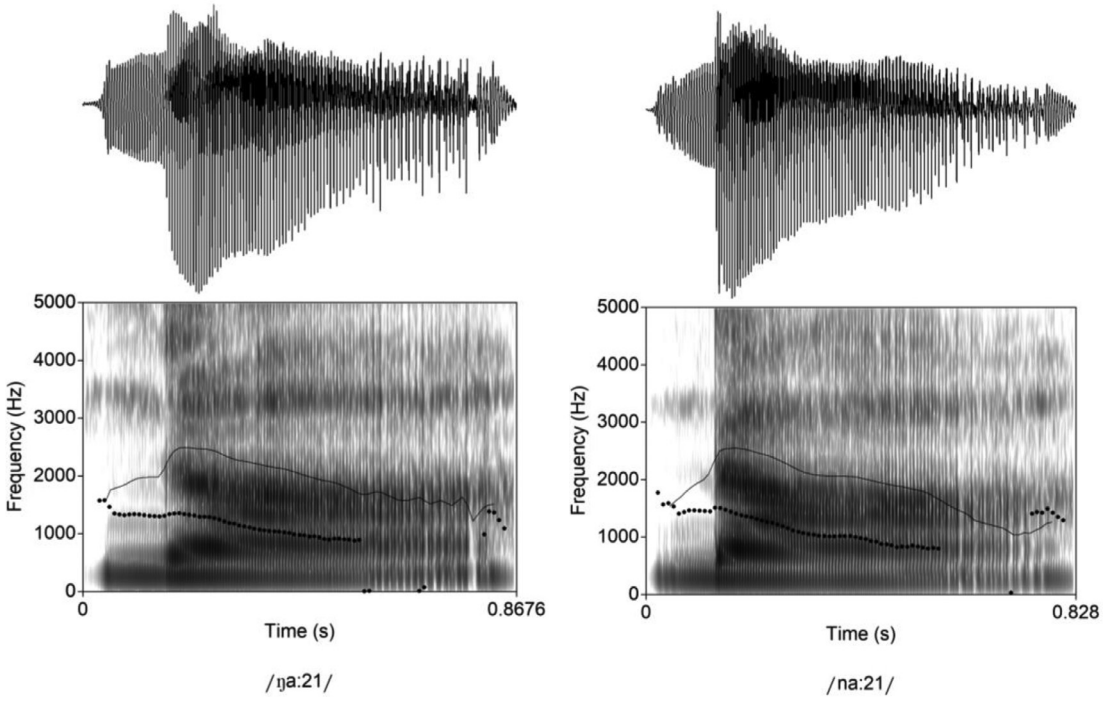

Fig. 6. CVV tokens /na:/ and /ŋa:/ with low tones.

examined suprasegmental discrimination using different vowels within trial, finding that the different-speaker-different-vowel condition was the most difficult. Furthermore, Wayland and Guion (2004) results indicated that different-speaker-same-vowel discrimination tests with five (three female, two male) talkers resulted in participants having little or no discrimination ability, even after training. Pilot testing with our stimuli revealed that participants were at ceiling in same-speaker-same-vowel tests and at floor in different-speaker-different-vowel tests. Therefore, the AX test used in the current study was different-speaker-same-vowel, with two talkers and the order of talkers in each trial was randomized.

As multiple talkers were used on each trial, issues may arise regarding the pitch height and range of the talkers. For example, if one talker tends to have higher pitch, then the height of a low tone may be comparable to the height of the other talker's mid tone. To address this issue, Table 1 provides the F0 ranges across tokens for both talkers in both tone types, and Fig. 8 presents the tone contours from the stimuli from Talker A and Talker B with normalized time. Table 1 and Fig. 8 illustrate two main differences between talkers. The F0 range of the mid tone differed because Talker B tended to drop the F0 at the end of the utterance and Talker A did not. Further, overall, F0 was higher for Talker B than

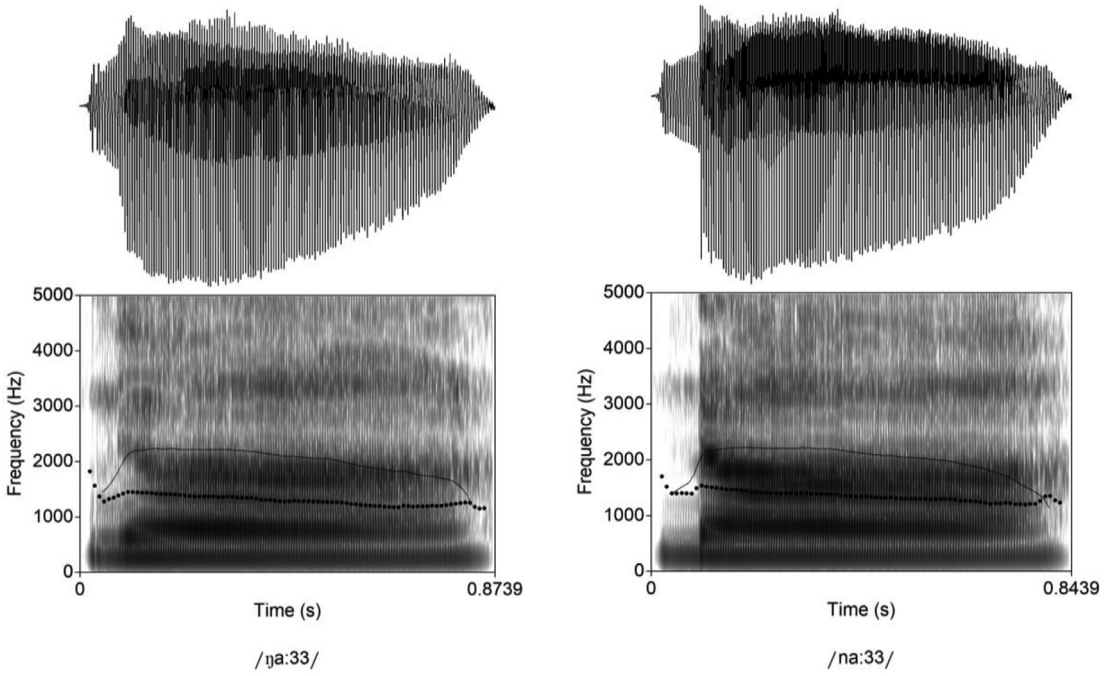

Fig. 7. CVV tokens /na:/ and /ŋa:/ with mid tones.

Table 1
F0 range, min, and max for Talker A and Talker B in mid and low tones across all tokens.

Talker	Tone	F0 range mean(sd)	F0 min	F0 max
A	mid	19.6 (7.1)	169.3	212.7
	low	50.4 (11.1)	128	202.9
B	mid	56.4 (7.3)	149.9	229.1
	low	50.9 (15.7)	152.5	238.3

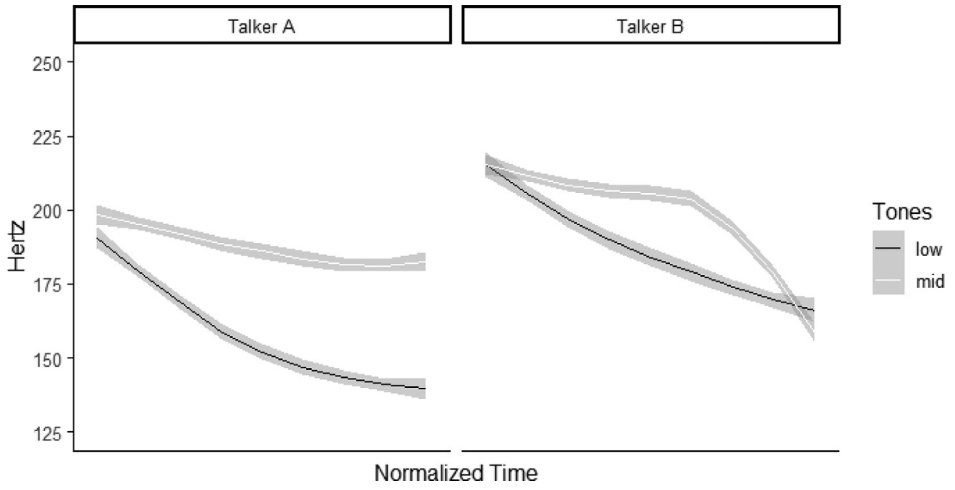

Fig. 8. Mean F0 contours and ± 1 standard error of the mean for mid and low tones across normalized time.

for Talker A. An immediate question is whether this resulted in differences in discrimination. It seems likely that the mid tone from Talker A would be perceived much more similarly to the low tone from Talker B than the opposite. Specifically, it is likely that discrimination accuracy would be lower on trials that contain Talker A mid tone and Talker B low tone than on trials with Talker A low tone and Talker B mid tone. This is measured and discussed in Section 5. It is important to note here that all participants heard the same trials, meaning that no participant had an advantage or disadvantage from hearing easier or harder trials.

As discussed, there were two sets of tokens. Each set contained 24 tokens composed from the stimulus types /kʰra:/, /pla:/, /na:/, /ŋa:/, /a:/, and hums with a mid tone and a low tone

for each talker. To make stimulus types equal, /aː/ and hums were duplicated for a total of 32 tokens for each set of tokens. Separate trial sets were created for each set of tokens. The two tokens in each trial were separated by a 1000 ms inter-stimulus interval. Every trial contained a single stimulus type and one token from each talker. Trials were counterbalanced on a variety of dimensions. Half of the trials had the same tone (i.e. no change trials) and half had different tones (i.e. change trials). Talker order (Talker A – Talker B, Talker B – Talker A) and tone order (low – mid, mid – low, mid – mid, low – low) were also counterbalanced, resulting in 32 trials for each syllable type. Trials were randomized into four blocks of 32 trials each. A final block of 64 trials included 16 trials from each syllable type. In total there were 192 trials for each participant.

Participants received a brief explanation of the difference between the tones used in the study, accompanied by examples. They were informed that they would hear two sounds on each trial and that the sounds would have a different tone or the same tone and they would need to decide if the tones were the same or different. The meaning of 'tone' was then explained to the participants. It was explained that some languages, along with consonants and vowels, have tones, which is where the pitch changes the meaning of the word and that the pitch can rise or fall or be high, medium, or low. Then an auditory example containing two productions of the same tone by the same Thai speaker was played and the participants were told that the tones were the same. Then another auditory example containing two productions of different tones by the same Thai speaker was played and the participants were told that the tones were different. The participants were also told that the utterances would come from different talkers and that the consonants would differ across trials. Finally, the participants were directed to attend to the tones and ignore other details. Then they did ten practice trials with feedback, consisting of five change and five no change trials. Practice trial feedback included whether they were correct or incorrect and what the tones were. After completing the practice trials, the participants completed the 192 trials of the main task. In the main task, feedback was not given. Finally, they completed a language background questionnaire.

4.4. Data analysis

Participant performance was analyzed via model comparisons of linear mixed-effects regression models from the lme4 package (Bates et al., 2015) within the R computing program (R Core Team, 2019). The independent variable was discrimination ability, measured through d' scores. A d' score is a measure of sensitivity to the signal that reduces decision bias. Fixed effects in the models include L1 background (native English or native Mandarin) and syllable type (CCVV, CVV, VV, or hums). Depending on the hypothesis under examination, syllable type was Helmert coded or simple effects coded. The interaction between these effects was also included in models. The maximal random effect structure that allowed the models to converge was used, which included random intercepts for participants. Model comparisons analyzed significance by testing whether a particular factor or interaction significantly contributed to the model fit.

5. Results

As discussed in Section 4.3, it is likely that discrimination accuracy would be lower on trials that contain Talker A mid tone and Talker B low tone than on trials with Talker A low tone and Talker B mid tone. To assess this question we measured accuracy on trials that contained these combinations. As expected, trials that contained Talker A mid tone and Talker B low tone were more difficult. Specifically, percent correct scores for native English listeners were lower on Talker A mid tone and Talker B low tone trials (35.6%, 20) than on Talker A low tone and Talker B mid tone trials (71.9%, 22). Percent correct scores for native Mandarin listeners also followed this pattern, being lower on Talker A mid tone and Talker B low tone trials (52.7%, 29.8) than on Talker A low tone and Talker B mid tone trials (82.7%, 19.7). These results suggest that there was a difference in trial difficulty attributable to talker variations in pitch. However, as mentioned in Section 4.3, all participants heard the same trials. Therefore, it is clear that the choice of talkers used in the stimuli set impacted the overall scores of the participants in the study, but as all participants heard the same stimuli, no group received an advantage from the stimuli.

Using model comparisons with linear regression models, we tested the hypothesis that experience with tone in a first language benefits listeners in novel tone discrimination. Native English speakers and native Mandarin speakers significantly differed on novel Thai tone discrimination (X^2 (1) = 15.5, $p < 0.001$). Native Mandarin speakers in this experiment were substantially better than native English speakers (see Fig. 9).

To see if the two groups differed as a function of syllable type, we tested the interaction between L1 background and syllable type, finding that discrimination across syllable types was not the same for both groups (X^2 (1) = 10.14, $p = 0.017$). To see where the groups differed, we compared discrimination scores across syllable types using Helmert coding. When comparing CCVV to CVV, VV, and hums, the difference between groups was significant (X^2 (1) = 8.47, $p = 0.004$), but there was no difference between groups when comparing CVV to VV and hums (X^2 (1) = 1.25, p = 0.26) or when comparing VV to hums (X^2 (1) = 1.25, $p = 0.26$). Further, the difference between CCVV and the other types was significantly greater for Native Mandarin speakers than for Native English speakers. We predicted that Native Mandarin speakers would do worse on CCVV than other types, but they actually did significantly better (see Fig. 9).

To further test the hypothesis that native Mandarin speakers would exhibit less accurate tone perception in CCVV syllables compared to CVV and VV syllables due to a lack of familiarity with complex onsets in the L1, we used a model comparison using Helmert coded linear regression models, which showed that d' scores for native Mandarin participants on CCVV syllables was significantly different from CVV and VV scores (X^2 (1) = 14.56, $p < 0.001$), while CVV scores were not different from VV scores (X^2 (1) = 2.02, $p = 0.15$). It was expected that CCVV scores would be significantly lower than CVV and VV scores, but CCVV scores were significantly higher than CVV and VV scores, as illustrated in Fig. 9.

We tested the hypothesis that native English speakers and native Mandarin speakers would exhibit less accurate tone perception in CVV syllables with /ŋ/ in the onset, which is

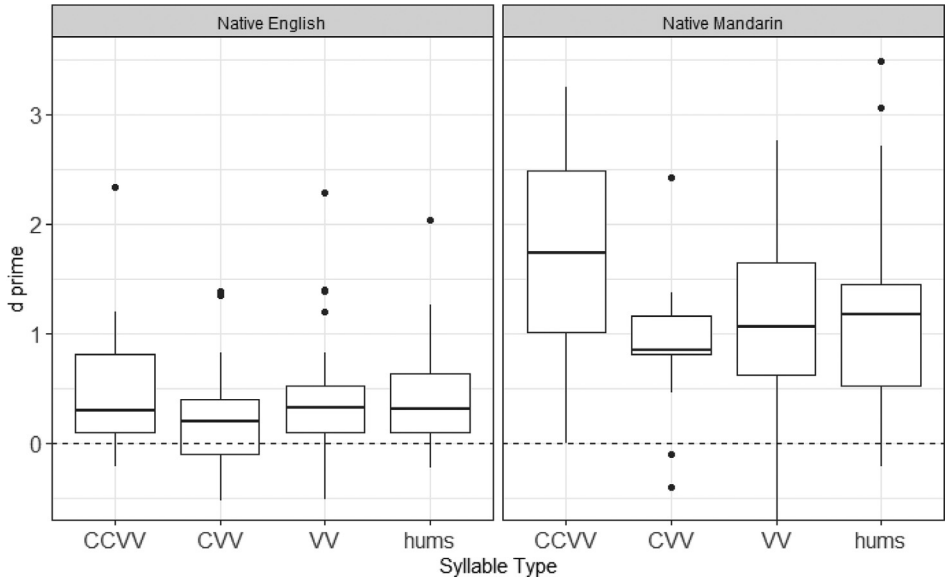

Fig. 9. Boxplot of Native English and Native Mandarin d' scores across syllable types.

phonotactically illegal in English and only allophonically allowed in Mandarin. Model comparisons using linear regression models showed that d' scores on /naː/ syllables were significantly different from /ŋaː/ syllables for both groups (X^2 (1) = 5.36, p = 0.02), and they showed that an interaction between L1 and CVV type was not significant (X^2 (1) = 1.097, p = 0.29). These results, illustrated in Fig. 10, suggest that having /ŋ/ in the onset negatively impacts tone perception, and this impact is not different for the two groups. However, while native English participants exhibited some discrimination ability between the tones in /naː/ trials, their discrimination scores for /ŋaː/ trials was at zero. We also tested the main effect of L1 background across CVV syllables, and results showed that d' scores for native Mandarin participants across CVV syllables was significantly different from native English participants' scores (X^2 (1) = 11.53, p < 0.001). Overall, native Mandarin participants were more accurate on CVV syllables.

6. Discussion

6.1. Summary of main findings

In the current study, we examined the effect of L1 background and phonotactic structure on novel tone discrimination. Specifically, we tested whether L1 background has an overall effect on novel tone perception across syllable types. Results suggested that native Mandarin participants were significantly better at discriminating between Thai mid and low tones than native English participants. We also examined the effect of syllable type on the discrimination of novel tones for each group to test our hypotheses. One hypothesis was that syllable structure would matter for novel tone discrimination. Specifically, one possibility was that phonotactic complexity would matter for novel tone discrimination, with discrimination being worse as complexity increased. However, results did not indicate a relationship between syllabic complexity and discrimination ability. Also, it was hypothesized that hums would differ from the other syllable types in predicting accuracy. Hums did not differ from CVV or VV but were worse than CCVV in predicting accuracy. It was also expected that the variability present in the onset of CCVV syllables would negatively impact discrimination accuracy. However, overall, participants did better at discriminating tones in CCVV syllables. Another hypothesis was that participants would do worse on phonotactic structures that were not legal in their L1. It was expected that Mandarin participants' tone discrimination ability would be hindered by CCVV structures. However, they actually performed significantly better on CCVV syllables compared to CVV, VV, or hums. Our stimuli set also permitted the examination of the effect of /ŋ/ onsets on novel tone discrimination. Results suggest that /ŋ/ onsets, which are illegal in English and only permitted allophonically in Mandarin, significantly hindered the discrimination of novel tones compared to /n/ onsets for both groups. For native English speakers the effect reduced discrimination ability to zero.

In the following sections we discuss results from the current study in relation to the hypothesis that syllable structure interacts with L1 background in predicting novel tone discrimination ability. Specifically, we address the effect of L1 background on novel tone discrimination and what the current results add to the ongoing discussion regarding the impact of first language experience with lexical tone on the ability to perceive and learn second language tone categories. We then discuss the effect of phonotactic structures on novel suprasegmental perception, addressing the long-held hypothesis that the inclusion of greater phonotactic complexity or novel phonotactic segments and structures in the stimulus set will hinder discrimination ability.

6.2. The effect of L1 background on novel tone perception

Results from the current study add to the growing body of evidence which suggests that predicting novel tone perception is more nuanced than a simple dichotomy between those with and without L1 tone experience. If we combine all trial types,

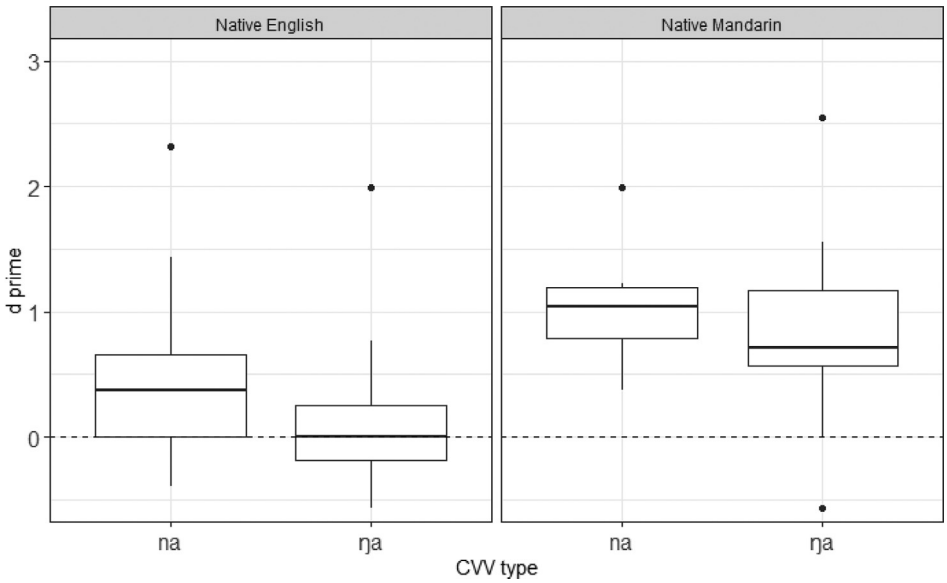

Fig. 10. Boxplot of native English and native Mandarin d' scores on the CVV types /na:/ and /ŋa:/.

we find that native Mandarin participants outperform native English participants. Thus, we could infer that L1 tone experience robustly accounts for novel tone perception ability, a conclusion frequently reached by many early tone perception studies. However, results from the current study, along with results from recent studies, suggest that novel tone discrimination ability is moderated by other factors. That is, an argument of improved tone perception just as a function of L1 background is overly simplistic for the real-world complexity of perception and acquisition of unfamiliar languages which use tone.

As discussed, overall discrimination scores in the current study follow results from studies suggesting that L1 tone experience robustly predicts novel tone perception and learning accuracy (Lee et al., 1996; Wang et al., 1999; Wayland & Guion, 2004; Wayland & Li, 2008; Qin & Mok, 2013). The explanation in these studies is typically that listeners with L1 tone experience are able to utilize their experience with lexical tone in their first language to more accurately perceive novel tones than those without L1 tone experience. The underlying hypothesis is that those with existing L1 tone categories are able to map the novel tones onto these L1 categories (see Reid et al., 2015; Chen et al., 2018, 2019, 2020), and this mapping results in a facilitatory effect during novel tone perception. Those without L1 tone categories, however, are unable to benefit from such mapping and thus have greater difficulty differentiating novel tones. It has been suggested that those without L1 tone experience may be able to utilize L1 intonational categories for novel tone category mapping (Francis et al., 2008), but more recent results indicate that L1 intonational categories are not as beneficial during novel tone mapping as L1 tone categories (Reid et al., 2015). We may be tempted to conclude that novel tone perception ability is determined solely from experience with L1 tone or intonation categories. However, a fixation on overall accuracy limits our perspective of novel tone perception.

Although L1 intonational experience may not be as beneficial for overall novel tone perception accuracy as L1 lexical tone experience, L1 suprasegmental experience in general impacts novel tone perception in a variety of ways. For example, depending on the listener's experience with L1 tone or L1 intonation the tones of the novel language may be easy or difficult to discriminate, identify, and learn (Francis et al., 2008; see also the body of work on the Perceptual Assimilation Model for Suprasegmentals (PAM-S) including So & Best, 2010). In the present study, we expand the understanding of which factors interact with L1 suprasegmental experience during novel tone perception by examining the impact of phonotactic structure on novel tone discrimination for those with and without L1 tone experience.

6.3. Phonotactic structure and novel tone discrimination

Phonotactic structure and familiarity impact novel tone discrimination in several ways and effects are modulated by the listener's L1 background. In the following sections we discuss these results in detail and analyze their impact on the current understanding of novel suprasegmental perception.

6.3.1. The effect of syllable structure on novel tone discrimination

Results from the current study suggest that greater phonotactic complexity across trials does not result in lowered discrimination ability across the board. That is, as complexity increases, novel tone discrimination ability does not decrease. Rather, in some circumstances complex onsets result in greater discrimination ability than tokens with more simple syllable structure. This was an unexpected result for several reasons.

First, phonetic and phonotactic stimuli variability has been included in some tone perception studies with results indicating phonetic variability and syllabic complexity have an inhibitory effect on novel tone discrimination. When there is greater segmental and syllabic variability within trial (Liu et al., 2011; Chen et al., 2019) and greater segmental complexity across trials (Qin & Mok, 2013), novel tone discrimination ability is reduced. Therefore, it was expected that greater segmental

complexity in the onset would inhibit tone discrimination, but that was not the case in the current study.

Second, most tone perception studies limit the syllabic complexity of stimuli to one syllable type, typically CVV, operating under the hypothesis that discrimination among temporally integrated segmental and suprasegmental features are difficult for beginning learners and naïve listeners to separate. Results from Liu et al. (2011) and Chen and Pederson (2017) support this hypothesis. Liu et al. (2011) found that beginning English learners of Mandarin performed the best in trials with simple onsets and consistent segments across tokens. When tokens within trial varied in segmental and syllabic composition, participants performed the worst at tone discrimination. Taken together, we reasoned that an inhibitory effect of syllabic complexity across trials may also occur, but an inhibitory effect was not found. Rather, participants discriminated suprasegmental information the best in trials with the greatest syllabic complexity.

The reason why syllabic complexity in the onset did not negatively impact novel tone perception across trials may be that onsets in the present study were not temporally integrated with tones to the extent necessary to direct attention away from tonal contrasts. In our stimuli, F0 patterns occurred predominantly over the duration of the nucleus, with the initial F0 ranges of the two tones being nearly identical. Thus, the segmental complexity introduced in the onsets of the stimuli did not occur simultaneously with the F0 information which was required to discriminate between the two tones. By contrast, in studies where syllabic complexity is greater across the nucleus and coda, novel tone discrimination is reduced (Qin & Mok, 2013; Liu et al., 2011). Therefore, it may be that syllabic complexity hinders novel suprasegmental discrimination proportionate to its level of temporal integration. Further, as both tones used in the study are identical in the initial part of each token, differences in the onset do not obscure contrasting F0 information that participants might rely on to discriminate tones.

It was also expected that syllable structure would interact with L1 background during novel suprasegmental discrimination. Results indicated that L1 background did interact with syllabic complexity on novel tone discrimination but not as expected. We expected that the lack of experience with complex onsets in the L1 would have a negative impact on L2 suprasegmental discrimination for native Mandarin participants. As complex onsets are not permitted in Mandarin, it was hypothesized that their presence would create an even greater distraction for native Mandarin participants than for native English participants. However, this was not the case. Native Mandarin participants discriminated between tones significantly better in CCVV syllables than in CVV syllables or VV syllables and this difference was significantly greater than native English participants.

One possible explanation for native Mandarin participants' accuracy on CCVV trials is that the native Mandarin participants in the current study have sufficient experience with languages that have complex onsets to not be hindered by CCVV structures. All participants were students recruited through the University of Oregon Psychology and Linguistics subject pool. Thus, all native Mandarin participants had met English proficiency requirements at a university level. Entrance at the University of Oregon requires scores of 88 on TOEFL, 7.0 on IELTS, and/or 120 on Duolingo English Test. Such experience with consonant clusters in English may have helped native Mandarin participants overcome any inhibitory effects complex onsets may have had. However, this explanation leads us to expect performance on CCVV syllables equivalent to CVV or VV syllables, but results from the current study indicated that participants did better on CCVV syllables than any other syllable type. Further, this explanation would not clarify why the increased discrimination ability in CCVV syllables over other types was greater for native Mandarin participants than for native English participants.

Another explanation for native Mandarin speakers' enhanced performance on CCVV syllables is that they perceived epenthetic vowels in CCVV syllables. Because complex onsets are illegal in Mandarin, these listeners may perceive CCVV syllables as CəCVV. Some studies suggest that naïve listeners listening to consonant clusters that are illegal in their L1 actually perceive a vowel between the consonants. Dupoux et al. (1999) found that Japanese listeners perceived epenthetic vowels in VCCV stimuli, which was an illegal phonotactic structure in their L1. Guan (2016) had Mandarin speakers transcribe Russian words containing consonant clusters and words without clusters, instructing them to assign tones to the syllables as well as segments. They frequently included epenthetic vowels between the consonants in a cluster. For stop + liquid clusters, the variety of cluster used in the present study, listeners primarily assigned a low tone to the epenthetic vowel. In the present study, if a low tone exemplar was initially activated by an epenthetic vowel, the native Mandarin participants may have experienced a perceptual benefit by being able to map that exemplar onto the Mandarin low tone. Studies show that Mandarin speakers consistently map the Thai low tone onto the Mandarin low tone (Reid et al., 2015; Chen et al., 2019). The activated exemplar in the epenthetic vowel may have then primed them to compare the following mid or low tone to the first activated exemplar. Thus, the task would not simply be discriminating between two novel tones. They would be discriminating between two words that were activating the same number or different number of exemplars, and this may be an easier task than discriminating between two novel tones. However, due to the lack of research on this area of speech perception, this hypothesis is speculative, and it may be unlikely given that the Mandarin speakers do have a relatively high proficiency in English, a language that uses CC clusters very robustly.

An alternate hypothesis, suggested by an anonymous reviewer, is that for Mandarin listeners, the only syllable type that is illegal is the CCVV type. Therefore, they were able to more easily assimilate all other tokens into their L1, perhaps hindering their perception. The CCVV type, however, could not be assimilated into any Mandarin category, potentially allowing for an increased discriminability in the tone categories for listeners. While it would be tempting to say that such an account would predict, however, that native English participants should then perform well on the /ŋ/ onset syllables, one could instead predict that the additional effort in processing an illegal syllable onset may disrupt their ability to process tones, which are already more challenging for them than the native Mandarin listeners.

One final account of the unexpected high scores on CCVV syllables by Mandarin listeners is that the greater segmental complexity of CCVV syllables resulted in a greater number of acoustic cues that could be used to distinguish tone categories. For example, in Section 3 we illustrated differences between hums and other tokens. Hums contained fewer cues that listeners could use to differentiate the tones. As discussed in Section 3, we tested the stimuli with Thai listeners to see if they were able to accurately identify the tones used. Hummed stimuli resulted in the lowest accuracy scores (82.5%). This may have been due to a lack of lexical meaning, but it could also have been due to a reduced number of cues that Thai listeners use to differentiate tone categories. By contrast, Thai listeners were able to identify the tones used in CCVV syllables with 96.3%, the highest of any token type. As Mandarin listeners were also able to perform the best in CCVV syllables and they did not have access to lexical meaning, it seems more likely that the CCVV stimuli contained more cues to aid in the differentiation of tone categories. If that is the case, then it suggests that Mandarin listeners were able to utilize the additional cues to achieve higher discrimination accuracy in CCVV syllables. Correspondingly, since native English speakers did not perform better on CCVV syllables than other syllable types, it would suggest that the native English participants in the study were not able to attend to the additional cues in CCVV syllables and use them above the other syllable types.

We also compared hummed tokens to other syllable structures. The results indicated that novel tone discrimination ability does not differ on hummed tokens compared to VV or CVV tokens for native English participants or native Mandarin participants. It was expected that hums may facilitate tone discrimination due to their low phonotactic complexity. Further, previous work has suggested that hums are processed differently than tokens with segments (Van Lancker & Fromkin, 1973). There is neuroscientific evidence that tones in words are processed as linguistic information, but tones in hums are processed differently, as non-linguistic pitch information (see Wang et al., 2004 for review). It has not been clear in previous work, however, whether differences in processing provide a benefit or detriment to suprasegmental perception. Results from the present study are inconclusive. It is also possible that there are other acoustic features that may have confounded the results in the hum perception condition. An acoustic analysis of the tokens revealed that voice quality differed between hums and tokens with segments. Tokens with segments often had creaky voice on low tones, and hums did not have creaky voice on low tones. It is often the case that a set of lexical tones in a language are not simply composed of variations in F0 but also can be composed of voice quality features (Gordon & Ladefoged, 2001; Yip, 2002; Kuang, 2013). As discussed in Gordon and Ladefoged (2001), creaky voice is often associated with lowered fundamental frequency. Stimuli in the present study followed this pattern. Creaky voice was often associated with low F0, meaning that low tones often had creaky voice while mid tones did not. This provided an acoustic cue that may have been exploited in trials with tokens containing segments, where low tones were often produced with creaky voice. For hums, however, both mid and low tones were consistently produced without creaky voice. This reduced the number of cues differentiating the tones in hums. Therefore, potential benefit from reduced phonotactic complexity in hums may have been negated by a reduced number of cues differentiating the tokens.

While our work directly investigates phonotactic structure, it is connected to work that investigates how syllable-tone co-occurrence may impact tone acquisition (Chan & Leung, 2020; Wiener & Ito, 2015, 2016; Wiener, Ito, & Speer, 2018, 2021). Taken together with the results of our present study, these data suggest that listeners co-process syllable and tone information and use both together in perception and acquisition of novel tones.

6.3.2. The effect of illegal segments in the onset on novel tone discrimination

In addition to investigations of how syllable structure might impact tone perception, the present study also examined the effect of illegal vs. legal onsets on novel tone discrimination. Specifically, the present study examined the effect of /ŋ/ onsets, finding that /ŋ/ onsets significantly hindered novel tone discrimination over /n/ onsets for both native English participants and native Mandarin participants. That is, the presence of an illegal segment in the onset had a negative impact on novel tone discrimination compared to legal onsets. One explanation is that perceptual difficulty emerges when attending to complex stimuli where multiple components are unfamiliar to the listener. That is, the greater the number of unfamiliar components, the greater the difficulty for the listener (see Liu et al., 2011). For native English speakers this means that syllabic complexity, which they are familiar with, creates less difficulty than /ŋ/ onsets, which are not phonemically permitted in English. The challenge that /ŋ/ onsets created for native English participants was so great that participants were unable to discriminate between novel tones when the trials included illegal onsets.

Native Mandarin speakers also experienced significantly reduced tone discrimination ability from /ŋ/ onsets, but their discrimination ability was not fully disrupted as it was for native English participants. This partial reduction may be due to the native Mandarin participants' familiarity with /ŋ/ as an allophone in onset position, which, in Mandarin, can fill the onset position when no other onsets are present (Duanmu, 2007). This partial reduction is in line with previous studies that suggest L1 allophonic experience with a segment is more beneficial for L2 perception than not having experience at all, but L1 allophonic experience is not as beneficial as L1 phonemic experience (Werker & Tees, 1984; Pegg & Werker, 1997; Whalen et al., 1997; Peperkamp et al., 2003). Results from the present study show that the degree and type of familiarity with segments in the onset of the target word has a proportional impact on the perception of novel suprasegmental features. These results confirm the hypothesis that a lack of familiarity with segments in the stimuli can have a negative effect on suprasegmental discrimination ability and build on this hypothesis by suggesting that the negative effect is modulated by the type of familiarity.

6.4. Interim summary

Native Mandarin participants did much better than native English participants overall and this is likely due to L1 tone

experience providing a greater benefit than L1 intonational experience during novel suprasegmental perception. However, an examination of other factors such as phonotactic features in the stimuli provides a more comprehensive understanding of novel suprasegmental perception.

Phonotactic complexity that is familiar to the listener may impact novel tone perception if it is temporally integrated with the target tones. In the present study complexity in the onset temporally preceded informative F0 information and may not have been fully integrated in a way that disrupted tone perception. Thus, native English participants, who had L1 experience with the CCVV, CVV, and VV syllable structures, did not vary across syllable types. Native Mandarin participants that lacked familiarity with complex onsets discriminated tones in CCVV trials differently from CVV, VV, or hums. They were more accurate on CCVV trials. We hypothesized that the most likely account for better performance on CCVV syllables by Mandarin listeners is that there were more salient acoustic cues in CCVV syllables and that native Mandarin listeners were able to attend to those cues while native English listeners were not able to do so. Overall, it was expected that CCVV syllables would negatively impact novel tone discrimination due to the variability present in the acoustic signal from production differences. However, the opposite may have occurred. The increased acoustic variability may have resulted in more cues to aid in the discrimination of the tone categories and listeners with L1 tonal experience were better able to use those cues.

The degree and type of familiarity with segments in the onset impacted novel tone discrimination for both groups. Native English participants demonstrated no ability to discriminate between tones when the onset was an illegal segment. For native Mandarin participants, discrimination was reduced compared to legal segments, but not to the same extent as native English listeners, likely due to their experience with /ŋ/ onsets as allophones in Mandarin.

A substantial body of work has demonstrated that a variety of factors impact perception and learning of novel segments and tones. The current work adds to that body of literature by demonstrating that, in the case of perception of novel tones, syllable structure impacts perception. Because tones are carried on syllables, it is logical that they would be co-processed, but this study is the first to demonstrate that CCV syllables and novel CV onsets impact how tones are perceived. Future work should investigate what other factors might impact perception and learning of tones, including how individual differences may interact with syllable structure and other properties of stimulus, task, and design (e.g., Perrachione et al., 2011).

6.5. Implications for novel tone category acquisition pedagogy

Teachers of tonal languages have to decide how they will teach the tone categories in the target language, making decisions about what is best for the students, both early in acquisition and over the course of learning, more broadly. Results from the current study provide some details that may be informative for teachers. Specifically, we demonstrate that there is an interaction between the phonotactic composition of carrier words and a student's language background during novel tone perception. Phonotactic complexity that the student is familiar with does not seem to impact tone perception; however, when the student is unfamiliar with the phonotactics in a word, variation in ability to discriminate between novel tones may occur. Students may have difficulty discriminating between tones when phonotactic complexity is temporally integrated with the target tones, as in the case of diphthongs or final consonants. Students may also have difficulty discriminating tones in words that include segments not found in their first language. With this knowledge, teachers can adjust their pedagogical methodologies to aid students in the initial acquisition of the target tone categories and also in the ongoing perception of tones in difficult phonotactic constructions.

Previous work has investigated how variability impacts learning. While this work has suggested that variability can improve learning, the question of how this variability ought to be introduced is an open question. That is, variability has both costs and benefits for learning and perception (e.g., Bradlow, 2008; Choi, Kou, & Perrachione, 2021; Lee & Baese-Berk, 2021; Perrachione et al., 2011). Therefore, it could be preferable for variability to be introduced piece-by-piece instead of introducing all variability simultaneously.

7. Conclusion

In the present study we examined whether phonotactic structure mattered for novel tone perception in two groups; native English speakers, who have experience with intonation but not lexical tone, and native Mandarin speakers, who have experience with lexical tone. We tested this question by examining novel tone discrimination ability across four syllable structures; CCVV, CVV, VV, and hums. Further, we compared novel tone discrimination ability in two types of onsets; /ŋ/ onsets, which are illegal in English but are allophones in Mandarin, and /n/ onsets which are legal in both English and Mandarin. It was predicted that native Mandarin participants would perform better on overall accuracy, that greater phonotactic complexity would negatively impact tone discrimination ability, that native Mandarin participants would perform worse on CCVV syllables, and that /ŋ/ onsets would negatively impact tone discrimination ability more for native English participants than for native Mandarin participants.

The results demonstrated that across syllable types, native Mandarin participants discriminated between novel tones better than native English participants, suggesting L1 tone experience provided a greater benefit than L1 intonational experience during the discrimination task. The results also demonstrated that greater phonotactic complexity in the onset did not impact novel tone discrimination for native English participants. It did impact discrimination for native Mandarin participants but not as predicted. Native Mandarin participants did substantially better on CCVV syllables than on the other syllable types regardless of the variability in the acoustic signal present in the initial consonant clusters of the CCVV syllables. Finally, the results demonstrated that /ŋ/ onsets negatively impacted novel tone discrimination ability for both groups as predicted. For native English participants the presence of /ŋ/ onsets, which are illegal in English phonotactics, resulted in no tone discrimination ability. Thus, we suggest that predicting novel tone perception ability is more complex than a simple dichotomy between L1 tone experience and L1 intonation

experience might suggest. The phonotactic structure of the target stimuli and the listener's experience with those structures can also impact novel tone perception.

CRediT authorship contribution statement

Jonathan Wright: Conceptualization, Methodology, Data curation, Software, Investigation, Writing – original draft. **Melissa Baese-Berk:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abramson, A. S. (1976). Thai tones as a reference system. *Tai Linguistics in Honor of Fang-Kuei Li*, 1–12.

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

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

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/lllt.17.07bes.

Bluhme, H., & Burr, R. (1971). An audio-visual display of pitch for teaching Chinese tones. *Studies in Linguistics, 22*, 51–57.

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [computer program] (2011). *Version, 5*(3), 74.

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English. *Phonology of Second Language Acquisition, 36*, 287–308.

Cao, R. (2012). *Perception of Mandarin Chinese Tone 2/Tone 3 and the role of creaky voice*. University of Florida.

Chan, R. K., & Leung, J. H. (2020). Why are lexical tones difficult to learn?: Insights from the incidental learning of tone-segment connections. *Studies in Second Language Acquisition, 42*(1), 33–59.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America, 128*(1), 456–465.

Chen, J., Best, C. T., & Antoniou, M. (2020). Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners. *Journal of Phonetics, 83* 101013.

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.

Chen, J., Best, C. T., Antoniou, M., & Kasisopa, B. (2018). Cross-language categorisation of monosyllabic Thai tones by Mandarin and Vietnamese speakers: L1 phonological and phonetic influences. Proceedings of the Seventeenth Australasian International Conference on Speech Science and Technology, 4-7 December 2018, Sydney, Australia, 169–172.

Chen, Y., & Pederson, E. (2017). Directing attention during perceptual training: A preliminary study of phonetic learning in Southern Min by Mandarin speakers. *Proc. Interspeech, 2017*, 1770–1774.

Choi, J. Y., Kou, R. S. N., & Perrachione, T. K. (2021). Distinct mechanisms for talker adaptation operate in parallel on different timescales. *Psychonomic Bulletin & Review*.

Davison, D. S. (1991). An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics, 78*, 50–57.

Duanmu, S. (2007). *The phonology of standard Chinese*. OUP Oxford.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1568.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(3968), 303–306.

Escudero, P. R. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Netherlands Graduate School of Linguistics.

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.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*(2), 268–294.

Gardner, R. (2001). *When listeners talk: Response tokens and listener stance* (Vol. 92). John Benjamins Publishing.

Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics, 29*(4), 383–406.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R.". *Neuropsychologia*, 317–323.

Guan, Q. (2016). Illusory vowels and illusory tones in the perception of consonant clusters by monolingual Chinese Mandarin speakers. 7th conference on Tone and Intonation in Europe (TIE 2016), University of Kent, Sep 2016, Canterbury, United Kingdom.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. *Language Experience in Second Language Speech Learning*, 57–77.

Jia, S., Tsang, Y. K., Huang, J., & Chen, H. C. (2013). Right hemisphere advantage in processing Cantonese level and contour tones: Evidence from dichotic listening. *Neuroscience Letters, 556*, 135–139.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica, 20*(2–4), 63–67.

Kuang, J. (2013). *Phonation in tonal contrasts* [PhD Thesis]. UCLA.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. *Handbook of Infant Perception, 2*, 275–382.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606–608.

Lee, D. Y., & Baese-Berk, M. M. (2021). Non-native English listeners' adaptation to native English speakers. *JASA Express Letters, 1*(10) 105201.

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research, 25*(5), 527–542.

Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a tonal language by attending to the tone: An in vivo experiment. *Language Learning, 61*(4), 1119–1141.

Maddieson, I. (2007). Issues of phonological complexity: Statistical analysis of the relationship between syllable structures, segment inventories, and tone contrasts. In M. J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 93–103). Oxford University Press.

Maddieson, I. (2013). Tone. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) The world atlas of language structures online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info, Accessed on 2021-10-30.)

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*(4), 369–390.

Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *The Journal of the Acoustical Society of America, 102*(6), 3742–3753.

Peperkamp, S., Pettinato, M., & Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th annual Boston University conference on language development* (pp. 650–661). Somerville, MA: Cascadilla Press.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America, 130*(1), 461–472.

Qin, Z., & Mok, P. (2013). Discrimination of Cantonese tones by speakers of tone and non-tone Languages. *Kansas Working Papers Linguistics, 34*, 1–25.

R Development Core Team. (2019). *R: A language and environment for statistical computing.*

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, & Psychophysics, 77*(2), 571–591.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*(4497), 947–949.

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.

Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association, 24*(3), 27–47.

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

Sun, S. H. (1998). *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese*. University of Hawaii Press.

Thepboriruk, K. (2009). Bangkok Thai tones revisited. *JSEALS, 3*, 86–105.

Van Lancker, D., & Fromkin, V. A. (1973). Hemispheric specialization for pitch and "tone": Evidence from Thai. *Journal of Phonetics, 1*(2), 101–109.

Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. A. (2004). *The role of linguistic experience in the hemispheric processing of lexical tone*.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America, 106*(6), 3649–3658.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning, 54*(4), 681–712.
Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics, 36*(2), 250–267. https://doi.org/10.1016/j.wocn.2007.06.004.
Werker, J. F. (1989). Becoming a native listener. *American Scientist, 77*(1), 54–59.
Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America, 75*(6), 1866–1878.
Whalen, D. H., Best, C. T., & Irwin, J. R. (1997). Lexical effects in the perception and production of American English/p/allophones. *Journal of Phonetics, 25*(4), 501–528.
Wiener, S., & Ito, K. (2015). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition and Neuroscience, 30*(9), 1048–1060.
Wiener, S., & Ito, K. (2016). Impoverished acoustic input triggers probability-based tone processing in mono-dialectal Mandarin listeners. *Journal of Phonetics, 56*, 38–51.
Wiener, S., Ito, K., & Speer, S. R. (2018). Early L2 spoken word recognition combines input-based and knowledge-based processing. *Language and Speech, 61*(4), 632–656.
Wiener, S., Ito, K., & Speer, S. R. (2021). Effects of multitalker input and instructional method on the dimension-based statistical learning of syllable-tone combinations: An eye-tracking study. *Studies in Second Language Acquisition*, 1–26.
Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28*(4), 565–585.
Yip, M. (2002). *Tone*. Cambridge University Press.
Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America, 136*(3), 1320–1333.