The influence of tone experience and native tone and intonation categories on nonnative tone learning

Jessica L. L. Chin, Mark Antoniou

The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia
jessica.chin@westernsydney.edu.au, m.antoniou@westernsydney.edu.au

ABSTRACT

For many languages, lexical tone is a linguistic feature which assigns meaning to words through variations in pitch. Learning tone as a new feature can be challenging, though training can improve learning outcomes for native tone and nontone listeners alike. Still, there is little consensus on whether native tone experience bolsters nonnative tone learning. This study examined the influence of the listener's native language on the learning of an artificial tone language. After five training sessions, nontonal Australian English, tonal Mandarin, and tonal Vietnamese listeners showed improvements in tone identification and word learning. Tone language listeners demonstrated greater performance in word learning, and Mandarin listeners outperformed the other groups in tone identification. Further investigation of tone identification patterns showed some tones were easier to perceive for all groups, though performance was also influenced by how closely the tones were mapped onto the listeners' native tone and intonation categories.

Keywords: tone training, tone perception, L1 experience

1. INTRODUCTION

Lexical tone is a linguistic feature that occurs in over half of the world's languages [1]. In a tone language, the meaning of words is distinguished by variations in pitch contours. These pitch contours—or tones—can vary in pitch height, direction, and trajectory. Pitch height is commonly numerically notated, where the lowest pitch value is 1 and the highest is 5 [2]. A level (or static) tone remains consistent in pitch height and direction, e.g., high level 55. A contour (or dynamic) tone can change in direction from onset to offset (e.g., rising 35), sometimes in a concave or convex shape (e.g., low dipping 214) [3]. In some languages, shorter "checked" tones occur only in syllables with a stop coda. Further, tone languages vary in the density of tones in their inventory, as well as in tone type. There is much variation between individual tones in a language, and between different languages.

For nontone language listeners, lexical tone is a novel feature to acquire, as these languages would only utilise pitch at the phrase/sentence level in the form of intonation [4]. Various studies show an advantage in nonnative tone perception for native listeners of a tone language over nontone language listeners [5, 6, 7], though some studies show no significant effect of tone experience [8, 9]. Features of individual tones in the target language, as well as tone presentation order, also strongly affect tone perception in native tone and nontone listeners [7]. In general, multi-session tone *training* has led to greater performance for all listeners at the cessation of training, regardless of tone experience. The influence of tone experience on learning performance is less clear [10]. In one study, listeners with a tonal L1 outperformed nontonal L1 listeners when learning nonnative tones [11]. There is also some evidence of a tone density advantage in one study comparing two listener groups with a different tonal L1 [12]. Sometimes, the tonal L1 learning advantage applies only to one task, yet performance is comparable between tonal and nontonal listeners in another (e.g., tone identification) [13]. Further, there are studies in the literature showing no advantages to nonnative tone learning for tonal L1 listeners over their nontonal counterparts. These outcomes were attributed to L1-influenced cue-weightings (towards pitch height and/or direction), and the relationship between L1 tone/intonation categories to the nonnative tones[13, 14]. In sum, a range of studies show variable effects of native tone language experience on nonnative tone learning. While there are advantages to having prior tone experience, this does not always apply across all perceptual tasks, and sometimes there are no advantages at all.

This study aimed to address the uncertain role of tone experience on nonnative tone learning. We examined the following effects: 1) native tone experience, 2) tone inventory density, and 3) similarities between native tone/intonation categories and nonnative tones. Native listeners of nontonal Australian English (AusE), tonal Mandarin, and tonal Vietnamese completed five sessions of tone training in an artificial language based on Hakka Chinese, a regional dialect spoken in Southern China. The tone inventory of Hakka and the tone/intonation inventories of the listener languages are summarised in Table 1.

We posit that prior tone experience would bolster nonnative tone learning for the native tone language listeners. If tone density predicts tone learning performance, the Vietnamese listeners would outperform the Mandarin. Conversely, if native tone/intonation categories are better predictors of nonnative tone identification, we would assume that the closer the nonnative tone maps onto the native category, the higher the identification accuracy. For example, Hakka T2 (11) maps closely to the low tone contours in Mandarin and Vietnamese. However, despite all listener groups having at least one category with a falling contour, Hakka's two falling tones (41 and 51) could still be too similar for nonnative listeners to distinguish consistently. The two falling tones begin and end at similar pitch heights and are predicted to be more challenging to learn overall.

Language	Tone/Intonation categories
Hakka [16], [17]	T1 (33), T2 (11), T3 (41), T4 (51) Checked: T5 (55), T6 (41)
English [18]	Flat pitch, rising question, falling statement, high falling exclamation
Mandarin [19]	T1 (55), T2 (35), T3 (214), T4 (51)
Vietnamese [20]	A1 (44), A2 (21), B1 (35), B2 (212), C1 (214) Checked: D1 (35), D2 (212)

Table 1: Tone and intonation categories of Hakka, English, Mandarin, and Vietnamese.

2. METHOD

2.1. Participants

Participants were native listeners of Australian English (n = 25; M_{age} = 24.46, SD = 8.2), Mandarin (n = 23; M_{age} = 26.11, SD = 5.4), and Vietnamese (n = 25; M_{age} = 24.53, SD = 6.9). None reported any hearing or neurological impairments, and all participants (except one Vietnamese listener tested remotely) passed an air conduction audiogram at 25 dB HL at 500, 1000, 2000 and 4000 Hz. Five English participants spoke another nontone language at home, and one participant studied Japanese in late adolescence. All Mandarin participants reported Mandarin as their dominant language. None spoke Hakka Chinese, though 14 reported knowledge of another Chinese dialect, and these contain their own lexical tone contrasts (n = 5 Cantonese, n = 3 Shanghainese, and six other dialects where n = 1). All but five Vietnamese participants spoke the Southern variant of the language; three participants spoke the Central variant and two spoke the Northern variant.

2.2. Stimuli

There were three experiment tasks: tone identification, tone word learning, and generalisation. The tokens in the tone identification task consisted of monophthongs /a/, /e̞/, /i/, /o/, /u/, and VC (checked) syllables /ak̚/, /ip̚/, /et̚/. The 16 tokens in tone word learning and generalisation consisted of CVC nonwords [fon], [leŋ], [nun], and checked syllable nonwords [wap̚] and [mip̚].

Auditory stimuli were produced by three female native speakers of Australian English, one for each task. The productions were recorded in a sound-attenuated booth and sampled at 44.1 kHz using a Shure SM10A cardioid microphone connected to a Roland Duo-Capture EX audio interface. All tokens were produced with a level tone, and were later superimposed with Hakka Chinese tones using the pitch-synchronous overlap and add function in Praat [21]. The Hakka tone onset and offset values were drawn from past analyses of three female native Hakka speakers [17].

Visual representations of Hakka's four tones and two checked tones were created for tone identification. In each image, an arrow depicted the height of the tone at onset and offset across five height levels. For tone word learning and generalisation, each of the 16 nonwords corresponded to a monochromatic image of a high-frequency word. These images were taken from a stimuli set used in previous research [21, 22]. Examples of the visual stimuli are provided in Fig. 1.

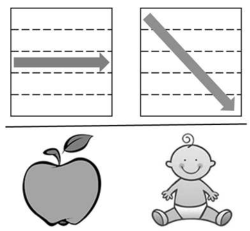

Figure 1: *Top row*: Visual representation of tones 1 (33) and 4 (51) in tone identification. *Bottom row*: Images corresponding to the nonwords [fon51] (apple) and [wap̚55] (baby) in tone word learning.

2.3. Procedure

Participants completed five sessions of training across separate days. Tasks were presented in a quiet testing space on a laptop PC running E-Prime 3.0. Stimuli were presented at 72 dB SPL over Sennheiser HD280 Pro headphones. In session 1, participants completed a demographics questionnaire, a tone identification pre-test, and tone word learning. In sessions 2 to 4, participants completed tone word learning only. In session 5, participants completed their final tone word learning session, a generalisation test, and a tone identification post-test. Three weeks after the cessation of training, participants were invited to complete a tone identification retention test. A total of 86% of participants returned for the retention test (68% AusE, 100% Mandarin, and 92% Vietnamese).

2.3.1. Tone identification

The tone identification task consisted of a familiarisation, practice, and test phase. Checked and unchecked tones were presented in separate blocks, and block presentation order was randomised for each participant. The tokens /a˥/ and /ak˥/ were used in familiarisation and practice, and were not included in the test phase. In familiarisation, participants were exposed to 3 repetitions of /a˥/ and /ak˥/. During practice, participants were presented with a sound, then instructed to select the tone they think they heard from the two visual response options on the screen (see Fig. 1 for examples). Feedback was provided after each response. In the test phase, participants were presented with the remaining stimuli in a similar format to the practice phase, except without feedback. Stimuli were presented at an interstimulus interval of 3 s; failure to respond within 3 s led to a missed response. A total of 104 trials were presented.

2.3.2. Tone word learning and generalisation

Tone word learning comprised of an exposure and test phase. During exposure, participants were presented with 4 repetitions of a minimal pair, with a unique image accompanying each sound. Participants were then presented with a sound and the two images as response options. Feedback was provided with every response. During the test phase, participants were presented with one sound at a time without feedback. All 16 images were provided as response options. A total of 96 test trials were presented.

The generalisation test was completed during the fifth session, directly after tone word learning. It was identical to the test phase of the tone word learning task, except with stimuli produced by a new talker.

3. RESULTS

Data were analysed using generalised linear mixed-effects models using the *glmer* function from the *lme4* package [24] in R [24, 25]. Post-hoc analyses were conducted with *emmeans* [27].

3.1. Tone identification

Mean identification accuracy is shown in Fig. 2. The model included the fixed factors Language (AusE, Mandarin, Vietnamese; Helmert coded) and Session (pre-test, post-test, retention test; repeated contrast coded), as well as random intercepts by participant and item. The Mandarin listeners outperformed both the AusE (β = 0.82, *SE* = 0.16, $p < .001$) and Vietnamese listeners (β = 0.81, *SE* = 0.19, $p < .001$). For all groups, performance improved between pre- and post-test (β = 0.22, *SE* = 0.02, $p < .001$), as well as between post-test and retention test (β = 0.14, *SE* = 0.03, $p < .001$).

We also examined tone identification patterns across listeners at post-test to determine whether identification was influenced by language background, individual tone characteristics, and/or the alternative response option provided in the task (see Table 2). Overall, the level tones were easier to identify than the falling tones, especially when the alternative response option was the other level tone. As predicted, the falling tones were most difficult for all listeners; in fact, the lowest identification scores (<75%) for the highest-performing Mandarin group were for these two tones.

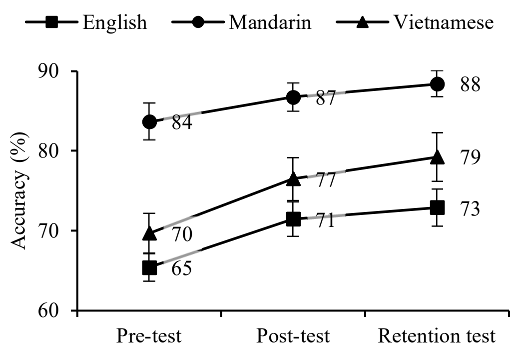

Figure 2: Mean tone identification accuracy by group at pre-, post- and retention test. Error bars depict SEM.

3.2. Tone word learning and generalisation

Fig. 3 shows mean accuracy for both word learning and generalisation. Word learning was analysed with a model that included the fixed factors Language (coded as above) and Session (1–5; repeated contrast coded), as well as random intercepts for participant and item. The model showed that both the Mandarin (β = 0.60, *SE* = 0.21, $p = .012$) and Vietnamese groups (β = 0.59, *SE* = 0.21, $p = .013$) outperformed their AusE counterparts. There was significant improvement in performance for all groups in subsequent sessions ($p < .001$). However, there was a

	Stimulus tone identification accuracy per alternative response option																	
	T1 (33)				T2 (11)				T3 (41)				T4 (51)				T5	T6
Alternative response option / Group	T2 (11)	T3 (41)	T4 (51)	T1 total	T1 (33)	T3 (41)	T4 (51)	T2 total	T1 (33)	T2 (11)	T4 (51)	T3 total	T1 (33)	T2 (11)	T3 (41)	T4 total	T6 (41)	T5 (55)
English	79%	66%	68%	**71%**	83%	75%	78%	**78%**	56%	68%	80%	**68%**	62%	82%	62%	**69%**	80%	76%
Mandarin	92%	97%	97%	**95%**	94%	79%	90%	**88%**	92%	81%	71%	**81%**	93%	94%	74%	**87%**	95%	91%
Vietnamese	89%	83%	79%	**84%**	84%	73%	82%	**80%**	69%	68%	78%	**71%**	72%	83%	72%	**76%**	82%	79%

Table 2: Tone identification (post-test) accuracy of stimulus tones when presented with different tones as alternative response options (e.g., English listeners identified T1 with 79% accuracy when T2 was the alternative response option). Overall tone identification accuracy of the stimulus tone is also provided.

significant interaction between the tonal groups and their performance from session 1 to 2 ($\beta = 0.10$, $SE = 0.05$, $p = .041$), suggesting that the Mandarin group improved more from session 1 to 2 than the Vietnamese group. The model used to analyse generalisation included the fixed factor Language (coded as above), and random intercepts for participant and item. Results showed that the Mandarin group outperformed the Vietnamese ($\beta = 0.53$, $SE = 0.24$, $p = .025$) and AusE groups ($\beta = 1.06$, $SE = 0.24$, $p < .001$).

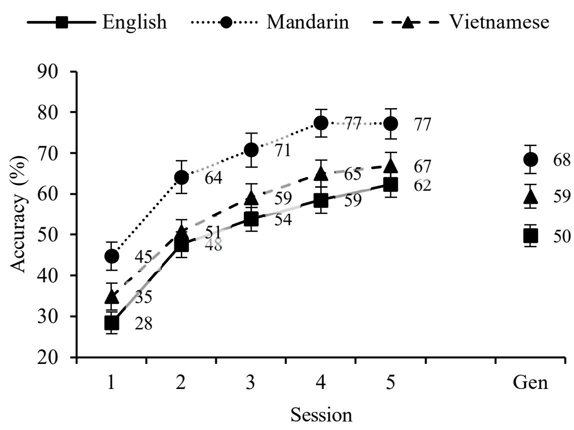

Figure 3: Mean tone word learning accuracy for all groups across all sessions and generalisation.

4. DISCUSSION

In the current study, listeners of varying language backgrounds completed multi-session tone training in an artificial language. All groups, regardless of level of tone experience, improved in tone identification and tone word learning at session 5. Further, all groups were able to retain these learning gains three weeks after the cessation of training. Native tone experience provides advantages to tone word learning, but not necessarily for tone identification: The Mandarin group outperformed both the AusE and Vietnamese groups in this task. Tone density also did not facilitate tone learning. Tone word learning requires listeners to engage in higher-level processes to link nonnative sounds to separate lexical items, much like learning words in a tone language. It would make sense, then, that tone language natives would be able to transfer this knowledge to a nonnative tone language. This finding falls in line with previous training studies demonstrating a tone language advantage [10, 11]. However, it is still uncertain whether native tone experience facilitates tone identification, or other factors bolstered the Mandarin listeners' performance. Studies have shown comparable performance in tone identification between tone and nontone speakers [12, 14], and have attributed this to factors such as cue weighting and L1 tone/intonation categories.

The tone identification patterns revealed a few tone-specific trends. Across listener groups, T2 (11) may be easier to identify due to its lower pitch and distance from the other Hakka tones at onset. Similarly, higher accuracy is observed for the stimulus tone T4 (51) when the alternative tone is T2 (11). Language-specific tone identification patterns were also present. The tone language listeners showed a smaller, but reliable, advantage in identifying T1 (33) than T2 (11), possibly due to T1's closeness to their native level tones relative to T2 (55 for Mandarin, 44 for Vietnamese). Despite the Mandarin group's overall higher identification accuracy, they misidentified the stimulus tone T3 (41) as T4 (51) more frequently than the other groups—a possible reason for this is interference from their own L1 falling tone (51). T3 also does not map closely enough to any categories in English (and neither falling tone maps closely to any Vietnamese category), which may explain the poorer identification accuracy for this tone.

In sum, while tone training can improve overall tone identification and word learning performance, tone- and language-specific identification patterns arise. There appear to be tones which are identified by all listeners with greater or poorer accuracy based on characteristics such as pitch height and direction. Further, the listener's native tone and/or intonation categories can influence nonnative tone identification, either to one's benefit or hindrance.

7. REFERENCES

[1] Yip, M. 2002. *Tone*. Cambridge University Press.

[2] Chao, Y. R. 1930. A system of tone letters. *Maître Phon.* 45, 24–27.

[3] Abramson, A. S. 1978. Static and dynamic acoustic cues in distinctive tones. *Lang. Speech* 21, 319–325.

[4] Ladd, D. R. 2008. *Intonational Phonology*. Cambridge University Press.

[5] Burnham, D. *et al.* 2015. Universality and language-specific experience in the perception of lexical tone and pitch. *Appl. Psycholinguist.* 36, 1459–1491.

[6] Schaefer, V., Darcy, I. 2014. Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Lab. Phonol.* 5, 489–522.

[7] Liu, L. *et al.* 2022. The tone atlas of perceptual discriminability and perceptual distance: Four tone languages and five language groups. *Brain Lang.* 229, 105106.

[8] Bent, T., Bradlow, A. R., Wright, B. A. 2006. The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 97–103.

[9] Tsukada, K., Xu, H. L., Rattanasone, N. X. 2015. The perception of Mandarin lexical tones by listeners from different linguistic backgrounds. *Chin. Second Lang. Res.* 4, 141–161.

[10] Antoniou, M., Chin, J. L. L. 2018. What can lexical tone training studies in adults tell us about tone processing in children? *Front. Psychol.* 9, 1–12.

[11] Wayland, R., Li, B. 2008. Effects of two training procedures in cross-language perception of tones. *J. Phon.* 36, 250–267.

[12] Tong, X., Tang, Y. C. 2016. Modulation of musical experience and prosodic complexity on lexical pitch learning. *Speech Prosody*, 217–221.

[13] Cooper, A., Wang, Y. 2012. The influence of linguistic and musical experience on Cantonese word learning. *J. Acoust. Soc. Am.* 131, 4756–4769.

[14] Wang, X. 2013. Perception of Mandarin tones: The effect of L1 background and training. *Mod. Lang. J.* 97, 144–160.

[15] Francis, A. L., Ciocca, V., Ma, L., Fenn, K. 2008. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *J. Phon.* 36, 268–294.

[16] Hashimoto, M. J. 2010. *The Hakka dialect: A linguistic study of its phonology, syntax and lexicon.* Cambridge University Press.

[17] Cheung, Y. M. 2011. Vowels and tones in Meixian Hakka: An acoustic and perceptual study. PhD dissertation, City University of Hong Kong, Hong Kong,

[18] Reid, A. *et al.* 2015. Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Atten. Percept. Psychophys.* 77, 571–591.

[19] Chen, C.-C. 2011. *Monographs on linguistics analysis: A synchronic phonology of Mandarin Chinese*. Walter de Gruyter.

[20] Brunelle, M. 2009. Tone perception in Northern and Southern Vietnamese. *J. Phon.* 37, 79–96.

[21] Boersma, P., Weenink, D. 2022. Praat: Doing phonetics by computer.

[22] Antoniou, M., Wong, P. C. M. 2016. Varying irrelevant phonetic features hinders learning of the feature being trained. *J. Acoust. Soc. Am.* 139, 271–278.

[23] Wong, P. C. M., Perrachione, T. K. 2007. Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.* 28, 565–585.

[24] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67, 1–48.

[25] R Core Team 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[26] RStudio Team 2020. RStudio: Integrated Development for R. RStudio, PBC., Boston, MA.

[27] Lenth, R. V. 2022. emmeans: Estimated Marginal Means, aka Least-Squares Means.