MANDARIN LISTENERS CAN LEARN NON-NATIVE LEXICAL TONES THROUGH DISTRIBUTIONAL LEARNING

Jia Hoong Ong, Denis Burnham & Paola Escudero

The MARCS Insitute, University of Western Sydney
jia.ong@uws.edu.au; denis.burnham@uws.edu.au; paola.escudero@uws.edu.au

ABSTRACT

In a previous study we have found that non-tone language speakers are able to form lexical tone categories through extracting frequency distribution in training, but only when attention is directed towards the distribution [12]. This study extends the distributional learning literature by investigating how tone language speakers' linguistic experience with tones affects their distributional learning of non-native lexical tones. Native Mandarin listeners were presented with a Thai lexical tone minimal pair distributed either unimodally (promoting formation of a single category), or bimodally (promoting two category formation). Assessment of performance in a discrimination task before and after exposure showed that the Bimodal Distribution group improved significantly from Pretest to Posttest whereas the Unimodal group did not. These results suggest that tone language speakers capitalise on their experience in using pitch phonemically to form the appropriate number of lexical tone categories based on the distribution that they hear.

Keywords: lexical tone; distributional learning; statistical learning; phonetic category.

1. INTRODUCTION

Distributional learning refers to the acquisition of categories by exposure to particular frequency distributions, which is thought to account for infants' acquisition of language-specific phonetic categories [20]. For example, given that Japanese has one liquid /r/ phoneme category, and English has two, the lateral /l/ and the rhotic /r/, speech input for liquids to Japanese infants would be concentrated around a single distributional peak, whereas native English infants will be exposed to two peaks along the same liquid continuum. Accordingly, in laboratory studies, distributional learning is said to occur when participants exposed to a bimodal distribution of a minimal pair show enhanced discrimination of the minimal pair whereas those exposed to a unimodal distribution show no such enhancement. Using this distributional learning paradigm, infants have been shown to acquire the appropriate number of phonetic categories based on the distribution structure of the input [11, 23].

With respect to distributional learning in adults, while it has been argued that the distributional learning effect is much reduced in adults compared with infants [18], evidence suggests that adults can indeed learn non-native sound contrasts through exposure to bimodal distributions [7, 8, 10, 14], suggesting that distributional learning may play a role in category formation in second-language acquisition. For example, native English language adults showed the predicted distributional learning effect of a Hindi minimal pair (/d/-/t/) when exposed to a bimodal but not a unimodal distribution of that minimal pair [10].

So far, distributional learning has mostly been investigated for consonants and vowels and lexical tones have been relatively neglected. In the 70% of world languages that are lexical tone or pitch-accent languages [22], pitch is phonemic such that a change of pitch height or contour on a particular syllable or pair of syllables can result in a different meaning. For example, in Mandarin the CV syllable /ma/ spoken with a high level tone (Tone 55) means 'mother' whereas it means 'scold' when spoken with a falling tone (Tone 51). The numbers here refer to Chao values, in which a relative scale of 1 (lowest) to 5 (highest) is used to represent pitch height and pitch contour across the duration of the syllable. The few distributional learning studies that have considered lexical tones suggest that after 12 months of age, learners do not seem to acquire lexical tone categories distributionally [9, 15] unless, as we found in a study of non-tone language adults, the learners' attention to the distribution is sustained throughout the exposure phase [12], analogously to the role of hyperarticulation in infant-direct speech [21]. In [12], non-tone language listeners (Australian English (AusE) listeners) in a bimodal distribution group outperformed those in a unimodal distribution group in discriminating a minimal tone pair manifested on a syllable different to that to which they were exposed. Furthermore, Posttest– Pretest difference scores showed that the bimodal group improved significantly on more test dimensions than the unimodal group, showing that there was actually *suppression* of learning due to exposure to a

distribution that hindered their discrimination of the lexical tone minimal pair.

In non-native speech perception studies it has been found that tone language listeners outperform non-tone language listeners in discriminating non-native lexical tones [e.g. 4, 19]. However, the extent to which tone language listeners can harness frequency distributions to learn non-native lexical tones compared with non-tone language listeners is not known. Here we investigate this issue by presenting unimodal and bimodal distributions of Thai lexical tones to native Mandarin listeners. It is possible that Mandarin listeners' extensive experience with lexical pitch is a double-edged sword in learning non-native lexical tones: on the one hand, they may be able to capitalise their experience of using pitch lexically to learn Thai lexical tones [13]; on the other, they may perform at ceiling and thus, not show any distributional learning at all [15].

2. METHOD

2.1. Participants

Thirty-six native Mandarin listeners (23 females) participated. Their ages ranged from 18 to 40 years old (M=24.83, SD=4.30). Some participants reported having minimal music training, though none had more than two years of training (≤ 0.5 year=2; 1 year=2; 1.5 years=2). All reported normal hearing. Participants were recruited from University of Western Sydney and University of New South Wales and were paid $15.

2.2. Stimuli

2.2.1. Distributional Learning Task

Four native Thai speakers (2 females) were recorded producing four Thai words: /k^{h}a33/, /k^{h}a241/, /na33/, and /na241/. A minimal pair was formed between two syllables differing only in tone (i.e., either /k^{h}a33/-/k^{h}a241/ or /na33/-/na241/). These minimal tone pairs were chosen because previous findings have shown that they are difficult for native Mandarin speakers to discriminate [3]. After initial inspection of the recorded tokens and matching for duration, four target minimal pairs were formed, each minimal pair produced by a different speaker. The four target minimal pairs differed in one or both of two dimensions: speaker's gender and syllable identity. To ensure that only the pitch contour differed between the members of each minimal pair of syllables, a base waveform of the same syllable spoken by the same speaker for each minimal tone pair was first chosen. Then, naturalistic pitch contours from the chosen tokens were imposed on those base waveforms for each minimal tone pair. The duration of syllables for the four minimal pairs ranged from 493ms to 832ms, but was kept constant within each minimal pair. These stimuli formed the reference exemplars used as test stimuli

Additionally, two target exemplars of each tone per minimal pair were generated using the method described above. This resulted in a total of 24 test stimuli (8 reference exemplars and 16 target exemplars). The target exemplars were used in test trials in which participants had to judge whether the target exemplar presented was similar to either of the two reference exemplars. The stimuli were normalised for amplitude at 70dB. For the training stimuli, the Male speaker-produced /na33/-/na241/ minimal pair (from the reference exemplars) was used to form an 8-step training continuum, spanning from Tone 33 (Token 1) to Token 8 (Token 241). The intermediate tokens (Tokens 2-7) were created by interpolating the pitch contours in equal steps.

As practice stimuli to familiarise the participants with the discrimination task, a sine wave tone and a sawtooth wave tone, both 440Hz and 800ms in duration, were generated using Praat [1].

2.2.2. Familiar Song Task

Given that tone language listeners are possibly more likely to possess absolute pitch (AP) [6], which may influence lexical tone perception [2] and given that the participants are not musically trained, we included a test of pitch memory performance, modelled after a previous study [17]. Based on a pilot study, we chose 40 popular English songs as the stimuli set for this task. The first 5s of each song was excised and duplicated. The duplications were randomly assigned to have their pitch raised or lowered either by one or two semitones. This resulted in four different sets (+1, +2, -1, -2 semitones), with each set having 10 songs. For the original excerpts, pitch was also transposed upward and then downward to the same degree as their duplications in order to remove any artificial artefacts due to digital manipulation.

2.3. Procedure

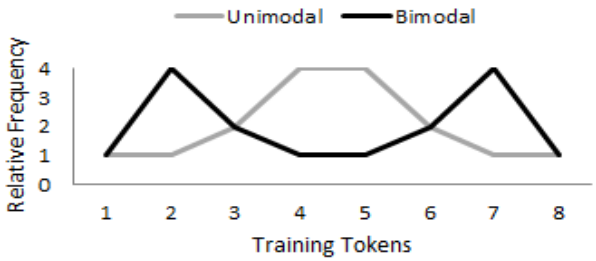
The experiment consisted of three tasks in the following order: (i) distributional learning task; (ii) familiar song task; and (iii) a language and music background questionnaire. The entire experiment took approximately 45 minutes to complete.

Distributional Learning Task: This task comprised three phases: Pretest, Training and Posttest. At Pretest and Posttest, the participants were required to perform an ABX discrimination

task, with the reference exemplars as A and B and the target exemplars as X. Each minimal pair was tested eight times, with order of test trials randomised. Participants were instructed that they had only 1s to respond, and there were no replacement trials for slow responses. Prior to the task, participants were given four practice trials to familiarise themselves with the ABX discrimination task using the practice stimuli.

In Training, participants were randomly assigned to one of two distribution conditions: Unimodal or Bimodal. They were instructed to listen to a sequence of sounds and to indicate on a paper response sheet whenever they heard a 'beep'. A total of 32 beeps were interspersed randomly within the training tokens. While both groups were exposed to the same total number of training tokens (256), the modal tokens heard by each group differed; the Unimodal group heard Tokens 4 and 5 most frequently, whereas the Bimodal group heard Tokens 2 and 7 most frequently (see Figure 1). Crucially, both conditions heard Token 1 and Token 8 (i.e., the reference exemplars of Male /na/ minimal pair used in the test trials) an equal number of times.

Figure 1: Distribution of Training Tokens heard by Unimodal (in gray) and Bimodal (in black) conditions.



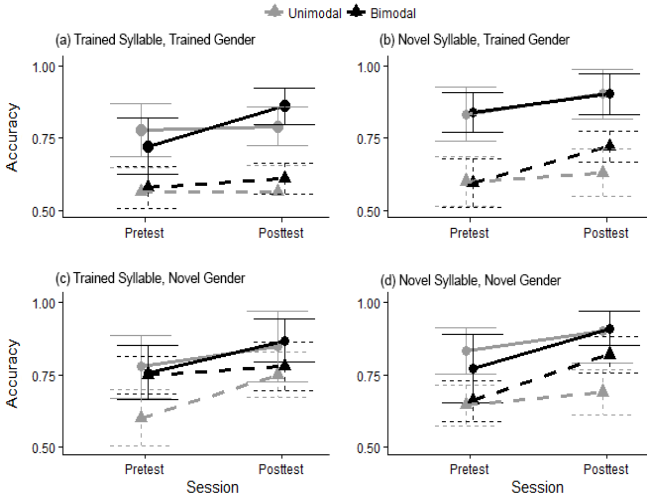
Familiar Song Task: Participants were first shown a song title and the artist who had performed the song, and asked whether they were familiar with the song. If they were familiar with the song, they heard two excerpts of that song, an original and a transposed version, with order of presentation counterbalanced, and they were required to indicate using a button press which excerpt was the original. If they were unfamiliar with the song, they moved to the next trial with no replacement trials. There were 40 trials in total, divided into two blocks: one with a transposition of ±1 semitone, the other with ±2 semitones, with presentation order within and between blocks randomised.

Language and Music Background Questionnaire: Participants were instructed to list (i) all the languages that they use and rate their ability on a 5-point scale in four different aspects: understanding, writing, speaking and reading and (ii) any musical training (either formal, private lessons or self-taught) that they may have had, the age of acquisition as well the length of training in years.

3. RESULTS

Prior to statistical analysis, participants' performance on the familiar song task was examined; it ranged from 33.33% to 77.78% correct (M=53.44%, SD=10.91%). No participant scored above 85% on average, so none met the AP possessor inclusion criterion [5]. An independent sample *t*-test was conducted to determine whether the two distribution conditions differed in their pitch memory performance. Quite unexpectedly, the Unimodal group (M=58.50%, SD=9.80%) had a significantly higher score on the familiar song task than the Bimodal group (M=48.40%, SD=9.80%; $t(34)=3.07$, $p=.004$). We will return to the implications of this later.

Figure 2: Discrimination performance on all 4 target minimal pairs by Distribution Condition: Unimodal (gray) and Bimodal (black). Solid lines = results of native Mandarin listeners in this study; dashed lines = native AusE listeners results reported in [12]. Error bars represent 95% confidence intervals.



A 2 x (2 x 2 x 2) Mixed ANOVA was conducted with between-subjects factor Distribution Condition (Unimodal vs. Bimodal) and within-subjects factors Session (Pretest vs. Posttest), Test Gender (Trained vs. Novel), and Test Syllable (Trained vs. Novel). There were significant main effects for Session ($F(1,34)=19.14$, $p<.001$) and Test Syllable ($F(1,34)=12.31$, $p<.001$), showing that (i) Posttest scores were higher than Pretest scores; and (ii) Novel Syllable (i.e., /k^{h}a/) test items were more easily discriminated than Trained Syllable (i.e., /na/) test items. Contrary to prediction, there was no significant Session x Distribution Condition interaction.

Following the analyses of previous studies [e.g., 7, 12], to determine whether the participants improved significantly from Prestest to Posttest, a series of one-sample *t*-tests were conducted on the difference scores on Trained and Novel Syllable (collapsing across Gender) as well as Trained and Novel Gender (collapsing across Syllable) test items for both Distribution Conditions. For the Unimodal group, none of the difference scores were significantly above zero, suggesting that there was no significant improvement whereas for the Bimodal group, the difference scores were all significantly above zero. The results of the one-sample *t*-tests seem counterintuitive; while there is no difference between the Bimodal and Unimodal groups in the Mixed ANOVA, the one-sample *t*-tests revealed significant improvement for the Bimodal group but not the Unimodal group. Upon inspecting the data more closely, we found that the difference scores for the Unimodal group are much more variable than those of the Bimodal group. Indeed, the coefficient of variation (CV) for the Unimodal group ranged from 2.04 to 7.00 whereas for the Bimodal group, the CV ranged from 1.00 to 1.38. This larger variance observed in the Unimodal group may have led to their insignificant results in the one-sample *t*-test.

4. DISCUSSION

The results indicate that the Bimodal group did not differ significantly from the Unimodal group at Posttest relative to Pretest, possibly due to the Unimodal group possessing better pitch memory despite random assignment of the participants. Nonetheless, the Unimodal group failed to show significant improvement from Pretest to Posttest, while those in the Bimodal group did. This suggests that the native Mandarin listeners showed a distributional learning effect of Thai lexical tones.

These results support the hierarchical inductive inferences model [13]. Since lexical tone is phonologically relevant in Mandarin, Mandarin speakers are able to acquire non-native lexical tones after just a brief six-minute exposure to a distribution. This is not to say that non-native tone language listeners are not able to do so, in fact the dashed line data in Figure 1 show that they do [12], but we predict that Mandarin listeners will show a *larger* distributional learning effect on lexical tones than non-native (e.g. AusE listeners) tone language listeners. Indeed, inspection of Figure 1 shows that while both the AusE listeners in the previous study [12] and Mandarin listeners in this study showed significant improvement from Pretest to Posttest in the Bimodal condition, the AusE participants improved only on three out of four test dimensions whereas the Mandarin participants improved on all four test dimensions. Furthermore, the AusE participants in the Unimodal condition showed no significant improvement except on one test dimension (Novel Gender) [12] while the Mandarin participants in the Unimodal condition did not show above-chance improvement on any of the test dimensions. Nonetheless, because there are far fewer Mandarin listeners in this (N=36) than AusE listeners in the previous study (N=50) [12], we are unable to conduct a direct quantitative comparison until more Mandarin listeners are recruited.

While we found that Mandarin listeners were able to benefit from their phonological experience in using pitch information to acquire non-native lexical tones distributionally, this study does not allow us to examine *how* they use this information, that is, whether the Mandarin listeners formed new tonetic categories or whether they shifted the category boundaries of their existing tonemic categories to assimilate the non-native Thai tones. In order to address this issue, future studies could include an assimilation task [e.g., 16] to establish how closely related Thai tones are to Mandarin tones before and after the training task. Future work could also address whether the advantage for tone language users is constrained to the same domain. For example, would native tone language listeners also show a clear distributional learning effect on *musical* pitch categories? Work is currently in progress in our laboratory to address this.

In sum, we found that native Mandarin listeners show a clear distributional learning effect for non-native Thai lexical tones. We argue that the Mandarin listeners' experience with their native tones transferred to learning new non-native tonetic categories (either by forming new tonetic categories or shifting existing tonemic category boundaries) based on the distribution of lexical tones that they heard. Preliminary comparison with a control study of AusE listeners suggests that tone language background enhances the learning of non-native lexical tone categories, over and above non-tone language experience, and further quantitative comparison of these groups is required. Future studies should also compare whether knowledge of phonemic pitch extends to learning non-linguistic pitch categories, such as musical pitch categories.

5. REFERENCES

[1] Boersma, P., Weenink, D. 2011. "Praat: Doing phonetics by computer". Version 5.3.42. http://www.praat.org.

[2] Burnham, D., Brooker, R., Reid, A. 2014. The effects of absolute pitch ability and musical training on lexical tone perception. *Psychol. Music*. doi:10.1177/0305735614546359

[3] Burnham, D. *et al.* 2014. Universality and language-specific experience in the perception of lexical tone and pitch. *Appl. Psycholinguist.* doi:doi:10.1017/S0142716414000496

[4] Burnham, D., Lau, S., Tam, H. & Schoknecht, C. 2001. Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers. *Proc. Auditory-Visual Speech Perception Conference 2001 (AVSP 2001)* 155–160.

[5] Deutsch, D., Henthorn, T., Dolson, M. 2004. Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Percept.* 21**,** 339–356.

[6] Deutsch, D., Henthorn, T., Marvin, E., Xu, H. 2006. Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *J. Acoust. Soc. Am.* 119, 719-722.

[7] Escudero, P., Benders, T., Wanrooij, K. 2011. Enhanced bimodal distributions facilitate the learning of second language vowels. *J. Acoust. Soc. Am.* 130, EL206–12.

[8] Hayes-Harb, R. 2007. Lexical and statistical evidence in the acquisition of second language phonemes. *Second Lang. Res.* 23, 65–94.

[9] Liu, L., Kager, R. 2014. Perception of tones by infants learning a non-tone language. *Cognition* 133, 385–394.

[10] Maye, J., Gerken, L. 2000. Learning phonemes without minimal pairs. *Proc. 24th BUCLD*, Boston, 522–533.

[11] Maye, J., Weiss, D. J., Aslin, R. N. 2008. Statistical phonetic learning in infants: Facilitation and feature generalization. *Dev. Sci.* 11, 122–134.

[12] Ong, J. H., Burnham, D., Escudero, P. Distributional learning of lexical tones: A comparison of attended vs. unattended listening. Submitted manuscript.

[13] Pajak, B., Levy, R. 2014. The role of abstraction in non-native speech perception. *J. Phon.* 46, 147–160.

[14] Perfors, A., Dunbar, D. 2010. Phonetic training makes word learning easier. *Proc. 32nd CogSci,* Portland, 1613–1618.

[15] Perfors, A., Ong, J. H. 2012. Musicians are better at learning non-native sound contrasts even in non-tonal languages. *Proc. 34th CogSci,* Sapporo, 839–844.

[16] Reid, A. *et al.* 2014. Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Atten. Percept. Psychophys.* doi:10.3758/s13414-014-0791-3

[17] Schellenberg, E. G., Trehub, S. E. 2003. Good pitch memory is widespread. *Psychol. Sci.* 14, 262–266.

[18] Wanrooij, K., Boersma, P., Zuijen, T. L. Van. 2014. Distributional vowel training is less effective for adults than for infants. A study using the mismatch response. *PLoS One* 9.

[19] Wayland, R. P., Guion, S. G. 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Lang. Learn.* 54, 681–712.

[20] Werker, J. F., Yeung, H. H., Yoshida, K. A. 2012. How do infants become experts at native-speech perception? *Curr. Dir. Psychol. Sci.* 21, 221–226.

[21] Xu Rattanasone, N., Burnham, D., Reilly, R. G. 2013. Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *J. Phon.* 41, 332–343.

[22] Yip, M. J. W. 2002. *Tone*. New York: Cambridge University Press.

[23] Yoshida, K. A., Pons, F., Maye, J., Werker, J. F. 2010. Distributional phonetic learning at 10 months of age. *Infancy* 15, 420–433.