Difficulty in learning L2 tones: Insights from the incidental learning of tone-segment mappings

Ricky Chan
Speech, Language and Cognition Laboratory, School of English, University of Hong Kong
rickykwc@hku.hk

ABSTRACT

Most previous studies on second language (L2) tone learning focused on explicit processes such as L2 tone identification/discrimination. However, the ability to identify/discriminate L2 tones does not entail the ability to encode pitch patterns as abstract tone categories at the word level, a pre-requisite for using them as lexical cues. This contribution reports an experiment on incidental learning of tone-segment mappings, which hinges on encoding pitch patterns as abstract tone categories at the word level. 80 Cantonese and English musicians and non-musicians were recruited. Results show that while the four subject groups distinguished the target tones similarly well perceptually, Cantonese speakers showed a learning effect of the target tone-segment mappings but English speakers did not, regardless of their musical background. This suggests that tone language speakers are able to implicitly categorize pitch contours as tone categories at the word level but non-tone language speakers cannot, regardless of their prior musical training.

Keywords: L2 tone learning, incidental learning, tone-segment mappings, music and speech

1. INTRODUCTION

Tone languages (e.g. Cantonese, Thai, Vietnamese) employ contrastive pitch patterns (i.e. lexical tones) to distinguish lexical meaning. Lexical tones are often reported to be difficult for L2 learners, especially for those whose native language (L1) is a non-tone language (e.g. English and French) (e.g. [5, 15, 16]). However, the nature of the difficulty involved is still far from clear. Most previous studies focused on explicit processes such as L2 tone identification or discrimination, and what factors—notably prior linguistic background and musical training—may contribute to better L2 tone perception and learning (e.g. [7, 10, 17]). However, the ability to perceive or discriminate different L2 tones explicitly does not entail the ability to encode pitch patterns as abstract tone categories at the word level, which is the pre-requisite of using abstract tone categories as lexical cues for word meaning contrast.

This paper is motivated by the hypothesis that, for learners whose native language is non-tonal, a major difficulty in learning novel lexical tones concerns repurposing pitch patterns from intonation cues to the formation of abstract tone categories at the word level. This hypothesis was tested with an experiment on the incidental learning (i.e. learning without intention) of tone-segment mappings. In tone languages, a given syllable may in principle carry any tones in their tone system. In some tone languages, however, the segments of a word may pose constraints on the possible tone a given word can carry at the lexical level ("tone-segment mappings"). For example, in many Chinese languages such as Cantonese and Hakka, entering tones only appear in words with a stop consonant such as /h/, /p/, /t/ or /k/ in the coda position [2, 9]. In Thai, the tone of a word is determined by a complex interplay among the initial consonant class, vowel length and syllable type. For instance, the mid tone and the rising tone only occur in syllables ending with a nasal stop, glide or an open vowel (see [13] for details). The learning of tone-segment mappings hinges on the perception and *encoding of both segments and abstract tone categories at the word level*, and thus tone-segment mappings are suitable learning targets for testing learners' ability to form abstract tone categories never before explored. The goal of the present study is to determine the effects of prior experience in the linguistic use of tones and musical training on the formation of novel abstract tone categories at the word level, as revealed by the incidental learning of tone-segment mappings.

2. METHOD

2.1. Participants

80 participants (20 Cantonese musicians, 20 Cantonese non-musicians, 20 English musicians and 20 English non-musicians[1]) were recruited.

2.2. Learning targets

The learning targets involved two artificial rules on the mappings between onset consonant and tone: i) words beginning with an aspirated stop (e.g. /p^h/, /t^h/ or /k^h/) carry a rising tone; ii) words beginning with an approximant (e.g. /l/, /w/ or /j/) carry a falling tone. Only monosyllabic words were used in the experiment, as tonal coarticulatory effect would have been a confound for the study of tone processing if polysyllabic words were used. To fully learn the rules above, participants had to be able to 1) distinguish

perceptually the two classes of consonants (aspirated stops vs. approximants) and the two lexical tones (rising or falling); and more importantly 2) form relevant segmental and abstract tone categories at the word level and pick up their connections. Since our focus was to determine whether participants could form novel tone categories at the word level, it was our intention to choose tones which should be easily distinguishable (rising vs. falling) at the phonetic-acoustic level, and onset consonants which are phonological natural classes in both English and Cantonese. No processing of meaning, which could be a confound in many previous studies on lexical tone processing, was required.

2.3. Stimuli and procedure

All stimuli used in the present study were monosyllabic nonce words generated by the Salika 2011 Thai speech synthesizer, and thus the phonetic realisations of the stimuli were based on Standard Thai. Thai was chosen in order to minimise the effects of prior linguistic knowledge for all the participants.
The learning of the target tone-segment mappings was assessed with a word learning task. Before that, an AX discrimination task was administered to test whether participants could perceptually distinguish the rising and falling tones in the learning targets.

2.3.1. AX discrimination task

The goal of the task was to test whether participants could distinguish the two tones (rising and falling) in the learning targets. The monosyllabic nonce words used in the task had either a CV or a CVV structure, which were concatenations of phonemes and tones:

- Onset: /s/, /h/ or /f/
- Nucleus: /i:/, /e:/, /ɯ:/, /ɤ:/, /u:/, /o:/, /ɔ:/, /i:a/, /u:a/ or /ɯ:a/
- Tone: rising (R) or falling (F)

The consonants /s/, /h/ and /f/ were chosen here in a bid to avoid overlapping with the phonological natural classes of the consonants involved in the learning targets. The concatenations above resulted in 60 different monosyllabic nonce words (3x10x2). In each trial, participants were auditorily presented with two monosyllabic words with an inter-stimulus interval of 500ms. They were instructed to indicate whether the two words have the same pitch pattern or not using a serial response box, and they were instructed to make their decision as quickly as possible. The two monosyllabic words in each trial were either AA pairs (same-tone pairs) and AB pairs (different-tone pairs). 60 AA pairs were formed by repeating the monosyllabic words (e.g. /fi:R/-/fi:R/; /hɤ:F/-/hɤ:F/). For the AB pairs, the two words shared the same segmental content and differed only in tone they carried (e.g. /su:aR/-/su:aF/. The order of the AB pairs was counterbalanced (i.e. both AB and BA pairs were used), resulting in 60 AB pairs.
A short practice was given at the beginning, followed by 120 trials in the AX discrimination task (60 AA pairs and 60 AB pairs). Both accuracy and reaction time data were collected. No feedback was given throughout the actual task. It was expected that the two target tones should be easy for all the participants to distinguish. If Cantonese and English learners performed similarly well in this task, a further question would be whether different participant groups would show differential success in learning the target tone-segment mappings.

2.3.2. Word learning task

The general design of the word learning task was adapted from Chan and Leung [3, 4]. The task consisted of a training phase and a testing phase.
Training phase. The goal of the training phase was to expose the participants to the target tone-segment mappings *incidentally* (i.e. without stating the target rules). 72 (4x6x3) monosyllabic nonce words used in the training phase were concatenations of the following phonemes:

- Onset: /p^h/, /t^h/, /l/ or /w/
- Nucleus: /i:/, /u:/, /o:/, /ɛ:/, /ɔ:/ or /a:/
- Coda: /m/, /n/ or /ŋ/

Words with an aspirated stop (i.e. /p^h/ and /t^h/) in the onset always carried a rising tone, whereas those with an approximant (i.e. /l/ and /w/) in the onset always carried a falling tone. By creating stimuli via concatenations of phonemes, the frequency of each phoneme in the nucleus and coda was the same for words with an aspirated stop onset and words with an approximant onset. This served to prevent participants from relating the tone a word carries with anything other than the nature of the onset consonant. Participants were told that they were going to learn words in an unknown language but were not told anything about the language. In each trial, participants were auditorily presented with a word. They were asked to listen to the word carefully and repeat it aloud as accurately as possible. To encourage the participants to pay attention to the pronunciation of the words, they were told that their voice would be recorded. No visual information on the spelling or meaning of the word was provided. Importantly, no explicit information about the connections between the initial consonant and tone type (i.e. the learning targets) was provided. Participants' pronunciation accuracy during the training phase was not the focus of the experiment and thus the production data were not analysed. The 72 nonsense words were randomly

presented and were repeated in four different blocks to form a total of 288 trials.

Testing phase (pronunciation judgment task). In each trial of the pronunciation judgment task, participants were auditorily presented with two words and they were asked to choose the one that "sounds better" to them (instead of "choose the correct one"). This encouraged the participants to use their intuition rather than explicit knowledge in their judgment [12]. Previous studies on incidental/implicit learning (e.g. [1, 6, 11]) have demonstrated that if participants possess abstract and potentially rule-like knowledge of the learning targets instead of merely memory of the training items, they should be able to transfer their knowledge to novel words/items. In this study we included two kinds of items: critical items and extension items [3, 4]. Sound pairs for the critical and extension items (see details below) differed only in terms of the tone they carried (e.g. /pʰɯːmR/ vs. /pʰɯːmF/). As such, it was when participants possessed knowledge of the target tone-segment mappings that they would show a preference for words which follow the target rules (e.g. /pʰɯːmR/ in the case above).

The critical items were concatenations of the following phonemes, resulting in 36 novel words:

- Onset: /pʰ/, /tʰ/, /l/ or /w/
- **Nucleus: /ɯː/, /eː/ or /ɤː/**
- Coda: /m/, /n/ or /ŋ/

In other words, the critical items differed from the items in the training phase *only in their vowel in the nucleus*. If participants had acquired abstract knowledge of the tone-segment mappings in the training phase rather than merely memorising the items encountered in the training phase, they should show a preference for novel words that follow the rules in the learning targets in the critical trials despite the change in the vowel.

On the other hand, the extension items were concatenations of the following phonemes, resulting in another 36 novel words. They differed from those in the training phase *only in their onset consonants*:

- **Onset: /kʰ/, /j/**
- Nucleus: /iː/, /uː/, /oː/, /ɛː/, /ɔː/ or /aː/
- Coda: /m/, /n/ or /ŋ/

Therefore, if participants had acquired abstract and potentially rule-like knowledge of the target patterns (e.g. if the onset consonant is an aspirated stop/approximant, the word should carry a rising/falling tone), they should be able to transfer their knowledge to words with a new onset consonant and show a preference for words which followed the learning targets in their judgment. All stimuli in the filler trials were words previously encountered in the training phase (50% of the all the trials [8]).

3. RESULTS

3.1 AX discrimination task

Figure 1 shows the average accuracy and log-reaction time (logRT) of the four groups for the AB pairs in the AX discrimination task. In general, the four groups achieved very high accuracy (92.4% to 95%) and their average reaction time was very close (2.55 to 2.73). Table 1 presents the results of analysis based on mixed-effects models; effect of any given factor was tested by likelihood ratio tests of a full model against a reduced model that excluded the effect to be tested. Models had maximal random slope structures as long as theoretically justified and can be converged in R. Musical background showed no significant effect on participants' accuracy or reaction time. Tonal experience showed a significant effect on participants' reaction time but not their accuracy. The interaction between Musical background and Tonal experience is non-significant for accuracy but significant for logRT. These suggest that participants' accuracy was mostly not affected by their prior musical training and tonal experience, but tonal experience has a relatively subtle facilitative effect for their reaction time, potentially at the automatized level. As expected, distinguishing a rising tone from and a falling tone as in the learning targets was generally easy for our participants.

Figure 1: Accuracy and logRT of the four groups for the AB pairs in the AX discrimination task. (C=Cantonese, E=English; error bar = 1 SE)

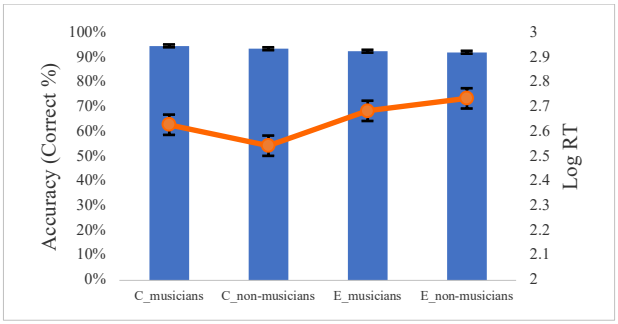

Table 1: Summary of the statistics of the mixed model comparisons for the factors of interest in the AX discrimination task.

Factor	Accuracy	LogRT
1) Musical background	$\chi^2(1) = 0.562$, $p = 0.454$	$\chi^2(1) = 0.498$, $p = 0.480$
2) Tonal experience	$\chi^2(1) = 0.960$, $p = 0.327$	$\chi^2(1) = 21.06$, $p << 0.001$*
3) Interaction of 1 and 2	$\chi^2(1) = 0.465$, $p = 0.495$	$\chi^2(1) = 99.08$, $p = <<0.001$*

3.2 Pronunciation judgement task

Figure 4 shows the average accuracy of the four groups for the critical items and extension items. For both critical items and extension items, most English participants performed at around chance level (50%) but most Cantonese participants performed considerably above chance level, regardless of their prior musical training. Analysis based on generalised linear mixed-effects models (see table 2) reveals that musical background had no significant effect on participants' accuracy on either critical items or extension items but tonal experience had a significant effect on their accuracy of both kinds of items.

Figure 2: Accuracy of the four groups for the critical items and extension items in the pronunciation judgment task (error bar = 1 SE)

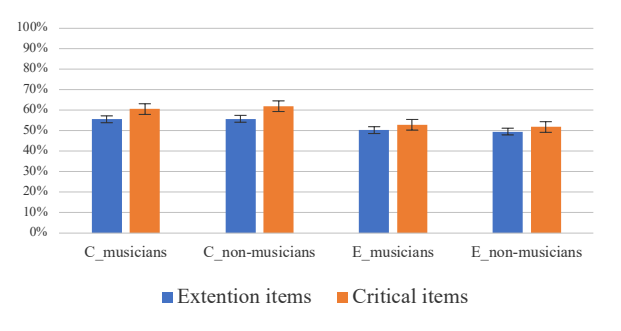

Table 2: Summary of the statistics of the mixed model comparisons for the factors of interest in the pronunciation judgment task.

Factor	Critical items	Extension items
1) Musical background	$\chi^2(1) = 0.001$, p = 0.975	$\chi^2(1) = 0.114$, $p = 0.915$
2) Tonal experience	$\chi^2(1) = 5.315$, $p = 0.0211$*	$\chi^2(1) = 5.105$, $p = 0.0239$*

Table 3 shows the 95% confidence intervals for the prediction of each participant group. For both critical items and extension items, the 95% confidence intervals of both Cantonese musicians and non-musicians do not contain the chance level value (50% accuracy), revealing that they performed significantly above chance and their knowledge of tone-segment mappings was abstract and potentially rule-like. However, the 95% confidence intervals for English musicians and non-musicians contain the chance level value, suggesting that they showed no learning effect of the target tone-segment mappings.

Table 3: Average accuracy (% correct) on and 95% confidence intervals (from upper bound to lower bound) of the critical items and extension items (chance level = 50%).

Group	Accuracy (critical items) (%)	Upper Bound	Lower Bound
C musicians	60.80	65.17	56.26
C non-musicians	62.22	66.53	57.70
E musicians	52.99	57.56	48.37
E non-musicians	51.85	56.44	47.24

Group	Accuracy (extension items) (%)	Upper Bound	Lower Bound
C musicians	55.65	59.89	51.33
C non-musicians	55.94	60.17	51.62
E musicians	50.28	54.60	45.96
E non-musicians	49.58	53.89	45.26

4. DISCUSSION

The present study sought to test the effects of prior musical training and tonal experience on the encoding of pitch patterns as abstract tone categories at the word level, as reflected in the incidental learning of tone-segment mappings. It was found that despite similar ability to distinguish tones perceptually, Cantonese learners could learn the target tone-segment mappings but English learners could not, potentially because English learners failed to repurpose pitch from intonational cues to forming tone categories at the word level [15]. Our study demonstrated that forming abstract tone categories at the word level may be difficult for English (and other non-tonal languages) speakers, which is a prerequisite of using abstract tone categories as lexical cues for contrasting word meaning. Further research in L2 tone learning should focus on how non-tonal language speakers may overcome such a processing bias and how to facilitate the encoding of pitch patterns as abstract tone categories at the word level by non-tonal language speakers. In addition, our findings also contribute to the long-standing effort in the study of the relationship between music and speech. Previous work has provided evidence for the overlap and separation of the musical and linguistic domains (see [10] for a review). However, a crucial question about the relationship between music and speech remain unexplored: i.e. whether prior musical training may facilitate the formation of abstract lexical tone categories. Our results show that prior musical training does not facilitate the encoding of pitch patterns as abstract tone categories at the word level. This highlights an area of the separation between the musical domain and the linguistic domain.

5. REFERENCES

[1] Altmann, G. T. M., Dienes, Z., Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 899–912.

[2] Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Walter de Gruyter.

[3] Chan, R. & Leung, J. (2014). Implicit learning of L2 word stress regularities. *Second Language Research, 30*(4), 463-484.

[4] Chan, R. & Leung, J (2018). Implicit Knowledge of L2 Lexical Stress Rules: Evidence from the Combined Use of Subjective and Objective Awareness Measures. *Applied Psycholinguistics*, *39*(1), 37-66.

[5] 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.

[6] Gomez, R. L. & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–35.

[7] Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*, 395–421.

[8] Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, *37*(1), 1-32.

[9] Lee, W. S., & Zee, E. (2009). Hakka Chinese. *Journal of the International Phonetic Association*, *39*(1), 107-111.

[10] Mok, P. P., & Zuo, D. (2012). The separation between music and speech: evidence from the perception of Cantonese tones. *The Journal of the Acoustical Society of America*, *132*(4), 2711-2720.

[11] Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford: Oxford University Press.

[12] Scott, R. & Dienes, Z. (2008) The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning Memory, and Cognition, 34*, 1264-88.

[13] Sladen, G. (2009). *Central Thai phonology*. Retrieved from

[14] http://www.thailanguage.com/resources/sladen-thai-phonology.pdf

[15] 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.

[16] 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.

[17] 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.

---

[1] Subjects with six years or more formal musical training and have played music/sang regularly in the past two years were classified as musicians; those with less than 2 years of casual musical experience and have not played music/sang regularly in the past 2 years were classified as non-musicians (based on [10]).