

Why are lexical tones difficult to learn? Insights from the incidental learning of tone-segment connections

Ricky KW Chan and Janny HC Leung

Speech, Language and Cognition Laboratory, School of English, University of Hong Kong

rickykwc@hku.hk, hiuchi@hku.hk

Acknowledgements

We would like to thank Professor Susan Gass, the editor, and two anonymous reviewers for their constructive feedback on our work. We would also like to thank Scarlett Hao, Alision Lam, Sula Ross, and Bruce Wang for their research assistance, and Patchanok Kitikanan for her advice on the stimuli. Part of the data collection took place when the first author worked in Lancaster University, UK. The financial support from Faculty of Arts and Social Sciences Research Fund, Lancaster University (project code: SZA1435) and Small Project Funding, University of Hong Kong (project code: 201409176014) is gratefully acknowledged.

Correspondence concerning this article should be addressed to Ricky Chan, Speech, Language and Cognition Laboratory, School of English, University of Hong Kong, Pokfulam Road, Hong Kong. E-mail: rickykwc@hku.hk

Abstract

L2 sounds present different kinds of challenges to learners at the phonetic, phonological, and lexical levels, but previous studies on L2 tone learning mostly focused on the phonetic and lexical levels. The present study employs an innovative technique to examine the role of prior tonal experience and musical training on forming novel abstract syllable-level tone categories. Eighty Cantonese and English musicians and nonmusicians completed two tasks: (a) AX tone discrimination and (b) incidental learning of artificial tone-segment connections (e.g., words beginning with an aspirated stop always carry a rising tone) with synthesized stimuli modelled on Thai. Although the four participant groups distinguished the target tones similarly well, Cantonese speakers showed abstract and implicit knowledge of the target tone-segment mappings after training but English speakers did not, regardless of their musical experience. This suggests that tone language experience, but not musical experience, is crucial for forming novel abstract syllable-level tone categories.

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

INTRODUCTION

Tone languages (e.g. Cantonese, Thai, Vietnamese) employ contrastive pitch¹ patterns at the syllable level (i.e. lexical tones) to distinguish word meaning. For instance, in Mandarin the syllable [ma] means “mother” when it carries a high level tone, but “scold” when it carries a high falling tone. Lexical tones are often reported to be difficult for second language (L2) learners, especially for those whose native language (L1) is a non-tone language (e.g. English and French) (e.g. Francis, Ciocca, Ma, & Fenn, 2008; So & Best, 2010; Wang, Spence, Jongman, & Sereno, 1999). However, the nature of the long-term 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. Hallé, Chang, & Best, 2004; Mok & Zuo, 2012; Wayland & Guion, 2004). 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 syllable level, which is the pre-requisite of using abstract tone categories as lexical cues for meaning contrast (i.e. the phonological level in a phonetic-phonological-lexical continuity in speech learning (e.g. Wong & Perrachione (2006)).

This paper is motivated by the hypothesis that, for learners whose native language is non-tonal, a major long-term difficulty in learning novel lexical tones concerns repurposing pitch patterns from intonation cues to the formation of abstract tone categories at the syllable level. This hypothesis was tested with an experiment on the incidental learning (i.e. learning without intention) of tone-segment connections (constraints by segmental composition on the possible tone a given syllable can carry), which hinges on the encoding of pitch patterns as abstract tone categories at the syllable level. On the other hand, the extent to which second

¹ While other phonetic cues such as duration, loudness and voice quality also play a role in tone perception and production, this paper will focus on pitch which is the primary cue of lexical tones.

language acquisition involves incidental learning or even implicit learning (i.e. learning without awareness) remains controversial (see Andringa and Rebuschat, 2015; Chan and Leung, 2018; Hulstijn and Ellis, 2005 for a review). An area of inquiry regarding incidental/implicit language learning lies in determining the scope of the adult capacity in acquiring knowledge of L2 language patterns without intention and/or awareness. Therefore, whether incidental exposure to tone-segment connections may lead to relevant abstract implicit knowledge is also an interesting theoretical question in language learning. The main goals of the present study are thus to 1) determine whether incidental exposure to tone-segment connections may lead to relevant implicit, abstract and potentially rule-like knowledge; and 2) 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 syllable level, and draw implications for L2 tone learning and the relationship between music and speech.

DIFFICULTY IN L2 TONE LEARNING

It is well known that unfamiliar L2 speech sounds may pose challenges for language learners, and L2 lexical tones are no exception. Lexical tones are often reported to be difficult for second language (L2) learners, especially for those whose native language (L1) is a non-tone language (e.g. English and French) (e.g. Francis et al., 2008; So & Best, 2010; Wang et al., 1999). Still, the nature of this difficulty is not well understood. In tone languages, pitch patterns function like vowels and consonants to contrast word meaning; on the other hand, in non-tone languages such as English, pitch patterns are mainly employed as intonational cues for contrasting pragmatic meaning or discourse types and functions (Wells, 2006). In terms of L2 tone processing and learning by non-tone language learners, three potential levels of difficulty in L2

tone learning can be identified (c.f. a phonetic-phonological-lexical continuity in speech learning (e.g. Wong & Perrachione, 2006):

- (1) perceiving and distinguishing tonal pitch patterns at the phonetic-acoustic level;
- (2) encoding pitch patterns as abstract tone categories at the syllable level; and
- (3) using abstract tone categories as lexical cues.

Most previous studies have addressed difficulty (1) by investigating explicit processes such as L2 tone perception or discrimination on monosyllables, and what factors may contribute to better L2 tone perception, notably i) whether learners' L1 is also a tone language; ii) proficiency in the target tone language; and iii) prior musical training. The key findings are summarized below.

Previous research has shown that tone-language speakers tend to outperform non-tonal language speakers in discriminating novel tones in general (e.g. Hallé et al., 2004; Leather, 1987; Lee & Nusbaum, 1993; Mok & Zuo, 2012; Repp & Lin, 1990; Stager & Downs, 1993; Wang, Jongman, & Sereno, 2001; Wayland & Guion, 2004), although perceptual success for individual tones depends crucially on the similarities of the tonal systems between the target language and the listeners' L1 as predicted by models of L2 speech perception such as the Perceptual Assimilation Model (e.g. Best, 1994; Francis et al., 2008; So & Best, 2010; Wayland & Guion, 2004). Generally speaking, non-tone language speakers appear to process fundamental frequency (acoustic correlate of pitch) differently from tone language speakers, both behaviourally and neurologically (Gandour et al., 2003; Hallé et al., 2004; Krishnan, Xu, Gandour & Cariani, 2005; Wang, Behne, Jongman & Sereno, 2004; Wong, 2002; Xu, Gandour, & Francis, 2006). Short-term training in tone identification has been shown to facilitate non-native tone identification for tone language speakers, but the results for non-tone language speakers are mixed (e.g. Francis et al., 2008; Wang et al., 1999; Wayland & Guion, 2004).

What about non-tonal speakers who have exposure to a tone language beyond one or a few experimental sessions? A relatively small body of work has found that experienced/advanced L2 learners of a tone language perform well or even close to native level in tone perception in monosyllables (Hao, 2012; Pelzl, Lau, Guo & DeKeyser, 2018; Wang et al., 1999; Zhang, 2011). This suggests that experienced/advanced L2 learners of a tone language appear to have little difficulty perceiving tones at the phonetic level.

Still, the ability to identify/distinguish different tone categories perceptually does not entail the ability to encode pitch patterns as abstract tone categories at the syllable level (difficulty (2)) and using abstract tone categories as lexical cues (difficulty (3)). Wong and Perrachione (2007) stresses the importance of examining the phonetic-phonological-lexical continuity in speech learning. Recent research in L2 speech learning has often shown a ‘discontinuity’ between phonetic level and the lexical level: the ability to identify/categorize L2 sounds phonetically does not *necessarily* predict success in lexical tasks (i.e. word learning in the target language) (e.g. Sebastián-Gallés & Díaz, 2012 at the segmental level). As for L2 tone processing and learning, while a few studies have demonstrated that in general non-tonal speakers can make significant progress on learning to associate tonal words and pictures instead of tonal labels only after a few experimental sessions, large individual differences exist (e.g. Bowles, Chang & Karuzis, 2016; Chandrasekaran, Sampath & Wong, 2010; Perrachione, Lee, Ha & Wong, 2011; Wong & Perrachione, 2007). Pelzl et al. (2018, p.4) pointed out that participants in these studies learnt “only a couple dozen words, all of which form minimal tone contrasts with other known words in the training vocabulary (e.g., *mā* vs. *má* vs. *mà*), thereby enhancing the salience of tones as a lexical feature. These studies thus are limited in their ability to reflect the realities of the much larger L2 Mandarin lexicon that experienced learners have acquired”. On the other hand, Malins and Joanisse (2010, 2012) used eye-tracking and ERP to demonstrate that the time course of lexical tone processing in native speakers is different from

that of acoustic-phonetic perception of tones. Wiener, Ito and Speer (2018) showed that L2 lexical tone processing draws on lexical information on which listeners cannot draw in a typical tone discrimination task. Also, while experienced L2 listeners have been found to perform comparably to native listeners on monosyllabic tone identification tasks (e.g., Lee, Tao & Bond, 2009), a few studies have demonstrated that even advanced L2 learners struggle with tone perception in disyllabic or polysyllabic stimuli (e.g. Hao, 2012; Pelzl et al., 2018; Zhang, 2011). For example, Pelzl et al. (2018) found that L1 English advanced learners of Mandarin showed near-native performance in a tone identification task for isolated words, but they struggled in a lexical decision task and a semantic judgment task which depended on the processing of tones on disyllabic words. As a speculation, the authors attributed such “discontinuity” to English learners’ inability to “repurpose” pitch patterns from intonational cues to lexical cues, but their study did not serve as a direct piece of evidence for this speculation in that, alongside other confounds (e.g. the small sample size and the wide age range of the participants), the difficulty arising from tonal coarticulation could have made individual tones in the disyllabic words in their study difficult to perceive (Chang & Bowles, 2015).

However, these studies on the ‘lexical learning’ of tones conflate lexical tone processing at the phonological level (abstract and categorisation) with the lexical level (processing of word meaning), and assumed that participants’ performance on lexical tasks directly reflects their ability to form phonological tonal contrasts. While successful ‘lexical learning’ of tones (e.g. being able to contrast words/meaning with tones) may entail successful ‘phonological learning’ of tones (i.e. possessing abstract tone categories), the absence of learning effects in relevant lexical or semantic tasks (despite success at the phonetic level) in those studies could have been due to difficulty in both the phonological and lexical levels, and their relative difficulty for learners remains unclear. In this paper we tease apart the phonological level from the phonetic

and lexical levels in tone learning and investigate learners' ability to establish abstract tone categories at the syllable level.

PRIOR MUSICAL TRAINING AND THE RELATIONSHIP BETWEEN MUSIC AND SPEECH

Music and speech are commonly believed to be closely related as both involve the use of pitch variation. Such overlap in the use of pitch raises interesting issues on the relationship between music and speech, specifically whether pitch processing is specific to the context in which it is learnt or whether pitch processing advantages span across different contexts. The study of music and speech has broader implications for the fundamental question in human cognition about how different "domains" in the mind may interact with one another. The majority of prior research focused on the perception of musical pitch and linguistic pitch by musicians and non-musicians. A number of studies have demonstrated that musicians generally outperform non-musicians in the identification of lexical tone (e.g. Gottfried, 2007; Lee and Hung, 2008; Mok & Zuo, 2012; Wayland, Herrera & Kaan, 2010), suggesting that prior musical training may facilitate lexical tone perception. There is also neurological evidence that musicians display enhanced performance in the processing of fundamental frequency (acoustic correlate of pitch) (e.g. Bidelman, Gandour & Krishnan, 2011; Lee, Lee & Shr, 2011; Wong et al., 2007). Further, Deutsch, Dooley, Henthorn and Head (2009) found that tone language speakers performed significantly better at a test of absolute pitch than non-tonal language speakers. A recent study has also revealed that Mandarin-speaking children performed significantly better at processing relative pitch than English-speaking children, even though both groups performed similarly on a control music task (timbre discrimination), suggesting that L1 tone language experience may enhance musical pitch perception in children because tone language experience drives attention

to pitch in non-linguistic contexts (Creel, Wang, Fu, Heyman & Lee, 2018). These findings suggest that there is overlap and interaction between the musical and linguistic domains, and that music and speech share very similar, if not the same, neural circuitry (Deutsch, Henthorn & Dolson, 2004) and pitch processing mechanisms (e.g. Bradley, 2012; Perrachione, Fedorenko, Vinke, Gibson & Dilley, 2013). The OPERA hypothesis was proposed to explain why prior musical training benefits the neural encoding of speech (see Patel, 2011 for details). Deutsch et al. (2009) even speculated that acquisition of absolute pitch by tone language speakers utilises the same mechanisms involved in the acquisition of a tone language.

However, there is also counterevidence suggesting the separation of musical and linguistic domains in the use of pitch. For example, Jiang, Hamm, Lim, Kirk and Yang (2010) and Nan, Sun and Peretz (2010) found no transfer effects from experience with linguistic use of pitch to musical pitch perception. Mok and Zuo (2012) found no facilitating effect of musical training on native lexical tone perception. Also, pitch range can vary considerably across individual talkers in their production of lexical tones, and thus success in acquiring L2 lexical tones hinges crucially on learners' ability to normalise and abstract tonal pitch contours across multiple tone tokens by different talkers and categorise them into different tone categories. In this regard, Wayland et al. (2010) found that although musicians performed better at pitch contour identification than non-musicians, their pitch contour abstraction and categorization ability was comparable to that of non-musicians, suggesting that prior musical training does not facilitate the abstraction and categorization of lexical tone categories.

Still, a crucial issue with regard to the relationship between music and speech remains unanswered: whether musical training may facilitate the formation of abstract tone categories at the syllable level by non-tonal language speakers, which is the prerequisite for the use of tone categories as lexical cues for contrasting word meaning. This question has not been directly addressed in the L2 tone learning literature and will be the main research question of

the present study. Findings will have implications for not only L2 tone learning, but also the relationship between music and speech.

WHY STUDY L2 TONE LEARNING VIA THE INCIDENTAL LEARNING OF TONE-SEGMENT CONNECTIONS

At the methodological level, most previous studies on L2 tone perception and processing employed tasks that involved the use of explicit knowledge related to tones (e.g. tone/word identification, lexical decision or semantic judgment based on minimal pairs that differ only in the tone). However, Pelzl et al. (2018) found that accurate explicit knowledge of tones and meanings for the target vocabulary items, as revealed in an explicit vocabulary knowledge test, was apparently enough for success in the tone identification task but not in the lexical decision task. They speculated that, for non-tonal L1 speakers, L2 tone may be difficult to acquire at the automatised/implicit level.

The review above suggests that 1) the use of non-explicit knowledge should be tested in L2 tone processing; and 2) we need a learning target that requires the encoding of abstract tone categories at the syllable level (phonological level) while avoiding confounds such as tonal coarticulation effects (phonetic level), the processing of meaning, and prior lexical knowledge in the target language (lexical-semantic level). The proposed study addresses these issues with a novel approach which involves the incidental learning of tone-segment connections.

In tone languages (e.g. Thai, Vietnamese, Cantonese), a given syllable may in principle carry different tones to contrast word meaning. In some tone languages, however, the segments (consonants and vowels) of a syllable may pose constraints on the possible tone a given syllable can carry (a kind of tonotactics/tonal phonotactics; “tone-segment connections” hereafter). For

example, in many Chinese languages such as Cantonese and Hakka, entering tones² only appear in syllables with a stop consonant such as /p/, /t/ or /k/ in the coda position (Bauer & Benedict, 1997; Lee & Zee, 2009; Sagart, 1999). In Thai, the tone of a syllable 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 Sladen, 2009 for details). The learning of tone-segment connections hinges on the perception and *encoding of both segments and abstract tone categories at the syllable level*, and thus tone-segment connections are suitable learning targets for testing learners' ability to form novel abstract tone categories.

The incidental learning paradigm was used. Incidental learning generally refers to the process by which information in the environment is picked up without conscious intention to learn (e.g. acquiring new words when reading a novel as a leisure activity) (Hulstijn, 2011). In contrast, explicit/intentional learning involves conscious intention and effort to learn and mainly results in conscious knowledge (Hulstijn, 2011). In laboratory settings, a typical incidental learning experiment usually involves subjects learning one aspect of the stimulus while paying attention to another. For instance, participants may pick up some grammar patterns when completing a meaning-focused task. On top of this, if the participants are unaware of the learning targets during the learning process, such situation is referred to as implicit learning (Williams, 2009). The present study focuses on the incidental learning mechanism in general.

It has been argued that the implicit/incidental learning mechanisms, often postulated as domain-general, play a fundamental role in various social behaviours and everyday situations,

² From a phonological point of view, there are “four tones” in Middle Chinese tonal system, namely, Ping (level), Shang (Rising), Qu (Departing) and Ru (Entering). Entering tone only occurs when syllables ending with stops, while the other three tones occur when syllables ending with vowel, semivowel and nasal (Sagart, 1999). Cantonese, for example, has three entering tones which have similar pitch heights and shapes as the three corresponding non-entering level tones but with shorter durations. They are often treated as allotones of the corresponding non-entering tones.

including first language acquisition (Reber, 1993). Still, the extent to which incidental/implicit learning mechanisms are involved in second language acquisition remains controversial (see Andringa and Rebuschat, 2015; Chan and Leung, 2018; Hulstijn and Ellis, 2005 for a review). A major line of research on incidental/implicit language learning thus lies in determining the scope of the adult capacity in acquiring knowledge of L2/novel language patterns without intention and/or awareness. Recently, research endeavours on incidental/implicit language learning of various aspects of language have surged (e.g. Rebuschat and Williams, 2012, Williams and Kuribara, 2008 on syntax; Leung and Williams, 2011, 2012, 2014 on form-meaning connections; Grey, Williams & Rebuschat, 2014, Rogers, Revesz & Rebuschat, 2016 on morphology). There is a relatively small but growing body of work with respect to phonological patterns. For instance, Dell, Reed, Adams and Meyer (2000) demonstrated implicit knowledge of novel artificial phonotactics ([f] always occurred as an onset and [s] as a coda), as revealed by speech errors, may be learnt through incidental exposure. Warker and Dell (2006) showed a similar incidental learning effect of second-order phonotactic constraints (e.g. if the vowel is [æ], [g] must occur as an onset and [k] as a coda, but if the vowel is [ɪ], [k] must occur as an onset and [g] as a coda). In the prosodic domain, recent studies have provided evidence that implicit knowledge of novel stress patterns (Chan and Leung, 2014, 2018; Graham and Williams, 2016) may be acquired through incidental exposure. This raises the possibility of acquiring implicit knowledge of other prosodic patterns incidentally/implicitly, and renders the possibility of acquiring implicit knowledge of tone-segment connections via incidental exposure a theoretically interesting research question. As discussed above, the learning of tone-segment connections hinges on the perception and encoding of both segmental and tone categories at the syllable level. This may pose additional challenges to learners, especially those whose L1 is a non-tone language. Therefore, another goal of this paper is to determine whether implicit knowledge of novel tone-segment connections may be learnt via

incidental exposure, and whether such learning may be constrained by learners' prior tonal experience and musical training.

A further question in implicit/incidental learning research concerns the nature of the resultant knowledge: specifically, whether incidental/implicit learning may lead to abstract implicit knowledge (see Chan and Leung, 2018 for a detailed discussion). Implicit knowledge refers to the unconscious knowledge which one is unaware of possessing, whilst explicit knowledge the conscious knowledge which one is aware of possessing and may be able to verbalize (Hulstijn & Ellis, 2005). Previous research has revealed a complex relationship between the conscious state of the learning process and that of the resultant knowledge. In other words, implicit/incidental learning does not necessarily lead to implicit knowledge, and explicit/intentional learning does not necessarily lead to explicit knowledge. For instance, implicit learning may first result in implicit knowledge. With continuous exposure to the stimuli, one may develop an 'insight' and relevant explicit knowledge of the underlying patterns in the stimuli. On the other hand, explicit learning may first lead to explicit knowledge but with enough practice, it may transform into implicit and automatised knowledge which influences behaviour without consciousness (Williams, 2009). In fact, both implicit and explicit knowledge may develop regardless of the learning process (e.g., Godfroid, 2016; Rebuschat and Williams, 2012; Rogers et al., 2016). Therefore, it is important to assess the conscious state of knowledge participants acquire after incidental exposure (Rebuschat, 2013). As for the abstractness of the resultant knowledge, controversy has centred on whether the resultant knowledge from incidental/implicit learning is abstract and potentially rule-based, or is merely based on memorised chunks/fragments or details of particular exemplars (see Chan and Leung, 2018 for a review). In the present study, participants' structural knowledge of tone-segment connections resulting from incidental exposure will be assessed via a source attribution task

(Dienes, 2008), and the abstractness of the resultant knowledge will be evaluated based on the transfer tests (e.g. Altmann, Dienes, & Goode, 1995) (see section 2.3 for details).

RESEARCH QUESTIONS

In sum, the research questions of the present study are as follows:

- 1) Can tone-segment connections be acquired through incidental exposure?
- 2) If so, is the resultant knowledge abstract and implicit?
- 3) What are the effects of prior musical training and tonal experience on the encoding of pitch patterns as abstract tone categories at the syllable level, as reflected in the learning of novel tone-segment connections?

METHODS

PARTICIPANTS

Four groups of participants (N=80) aged 17-30 were recruited in the study (20 Cantonese musicians, 20 Cantonese non-musicians, 20 English musicians, and 20 English non-musicians) and their background information are summarised in Table 1. All Cantonese learners spoke English and Mandarin as an L2, ranging from intermediate to advanced level for English and from beginner to advanced level for Mandarin based on self-report. Some L1 English learners spoke other non-tonal languages—including German, Spanish, French, Hungarian, Dutch, Welsh, Swedish, and Persian, ranging from beginner to advanced level based on self-report, but none of the English learners reported any knowledge of any tone language.

Musicians and non-musicians were defined as follows based on Mok and Zuo (2012):

- Musicians: participants with 6 or more years of formal training in singing or any instrument who have played music regularly in the past 2 years;

- Non-musicians: participants with no more than 2 years of casual musical experience/training who have not played music regularly in the past 2 years

People with musical experience between the two categories were not recruited for this study as their musical background was too ambiguous.

TABLE 1. Background information of each participant group.

Participant Group	Mean age	Gender breakdown	Mean years of musical training
1) L1 Cantonese musicians	21.3	6M, 14F	8.4
2) L1 Cantonese non-musicians	22.6	11M, 9F	0.59
3) L1 English musicians who had not learnt any tone language	21.8	8M, 12F	7.7
4) L1 English non-musicians who had not learnt any tone language	23.1	9M, 11F	0.71

LEARNING TARGETS

The learning targets involved two artificial rules on the mappings between onset consonant and tone as illustrated below:

1. Words beginning with an aspirated stop (e.g. /p^h/, /t^h/ or /k^h/) always carry a rising tone (R) (e.g. /p^ho:mR/, /t^hɔ:ŋR/, /k^ha:nR/).
2. Words beginning with an approximant (e.g. /l/, /w/ or /j/) always carry a falling tone (F) (e.g. /lo:mF/, /wɔ:ŋF/, /ja:nF/).

Only monosyllabic words were used in the experiment, as tonal coarticulatory effect would be a confound for the study of tone processing if polysyllabic words are 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 syllable level and pick up their connections. Since our focus was to determine whether participants could form novel tone categories at the syllable level, it was our intention to choose tones which should be easily distinguishable (rising vs. falling) at the phonetic-acoustic level, and used phonological natural classes that are present in both English and Cantonese. No processing of meaning, which was a confound in many previous studies on lexical tone processing, was required.

Participants' prior linguistic knowledge may play a role in the learning of the target patterns. At the segmental level, /p^h/, /t^h/ and /k^h/ and /l/, /w/ and /j/ are phonological natural classes of aspirated stops and approximants respectively in both Cantonese and English. As such, all participants in the experiment have similar starting points for the segmental part of the learning targets³. At the prosodic level, Cantonese contrasts six phonemic tones: high level[55], high rising[25], mid level[33], low falling[21], low rising[23], and low level[22] (Bauer, 1997). The Cantonese learners also spoke Mandarin as an L2 which contrasts four phonemic tones on monosyllables: T1[55] high tone, T2[25] rising tone, T3[214] dipping/low tone, and T4[51] falling tone (Norman, 1988). The rich pitch contrasts in the Cantonese and Mandarin tone systems, and in particular the presence of both rising and falling tones, may facilitate their perception of novel rising and falling tones in the learning targets (e.g. they may perceptually assimilate the target tones to the Cantonese and/or Mandarin counterparts as predicted by the Perceptual Assimilation Model (So & Best, 2010). Also, native Cantonese speakers should be able to encode novel abstract tone categories at the lexical level as tones are used contrastively in Cantonese at the syllable level. On the other hand, native English speakers should have little problem distinguishing between a rising tone and a falling tone as

³ Pairs of aspirated and unaspirated stops (e.g. /p/ and /p^h/) are allophones in English but are different phonemes in Cantonese. However, this should not lead to any difference in learning the target tone-segment connections by the four groups of learners since in the present study, stops only appeared at a fixed location (i.e. syllable onset in monosyllabic words) and were not preceded by /s/. Under these conditions, only phonetically aspirated, and not unaspirated, voiceless stops occur in English.

the two pitch contours are frequently used in English intonation (Wells, 2006). They could potentially assimilate the novel tones in the learning targets to their intonational rise and fall of the English tonal system (So & Best, 2010). However, although contrastive pitch patterns may also be found on monosyllabic words in English (e.g. the single word “me” may carry a rising tone or a falling tone to convey uncertainty and definiteness respectively), in English intonation typically operates at the utterance level involving more than one syllable and English does not contrast word meaning based on syllable-based pitch variation. Native English speakers may not be able to encode abstract tone categories at the syllable level and thus they may fail to learn the target tone-segment connections.

STIMULI AND PROCEDURE

All stimuli used in the present study were monosyllabic nonce words generated by the Salika 2011 Thai speech synthesizer⁴, 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. None of the participants reported any knowledge of Thai. The experiment was administered on a computer using E-prime.

The learning of the target tone-segment connections was assessed with a word learning task. Before that, an AX discrimination task was administered to test whether participants could perceptually distinguish the two tones in the learning targets (rising and falling).

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 listed below:

⁴ By PPA Innovation Co., Ltd. (see http://www.ppainnovation.com/salika/homeedition_en.html for details). Synthesised stimuli instead of recordings by a native speaker were used in the experiment because a native speaker is likely to vary in their fluency when trying to pronounce nonce words.

- Onset: /s/, /h/ or /f/
- Nucleus: /i:/, /e:/, /ʊ:/, /ɜ:/, /u:/, /o:/, /ɔ:/, /i:a/, /u:a/ or /ʊ: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 (ISI) 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/; /hɜ:F/-/hɜ: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 connections.

Word Learning Task

The general design of the word learning task was adapted from Chan and Leung (2014, 2018). All the monosyllabic words used in this task had a CVC structure. The task consisted of two phases: the training phase and the testing phase.

Training phase. The goal of the training phase was to expose the participants to the target tone-segment connections *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.

The design of the training phase was in line with the “noticing” hypothesis in implicit/incidental learning (Schmidt, 1990, 2001, 2010), which stipulates that “noticing” (conscious registration of attended input) but not “understanding” (a higher level of awareness that involves generalisations across instances, such as knowledge of rules and metalinguistic awareness) is required for the incidental/implicit learning of language patterns. In the training phase of the present study, incidental/implicit learning of the target rules was hypothesised to depend on attention to the pronunciations of the segments and tones⁵ through practice, but generalisations of tone-segment connections beyond individual training items and the emergence of relevant abstract implicit knowledge could take place via a basic associative learning mechanism in humans which automatically extracts patterns across instances, leading to a non-verbalisable and intuitive form of knowledge (Schmidt, 2010).

Testing phase. The testing phase consisted of a pronunciation judgement task and a source attribution 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 (Scott & Dienes, 2008).

Previous studies on incidental/implicit learning (e.g. Altmann et al., 1995; Gomez and Gerken, 1999; Reber, 1993) 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

⁵ By performing the AX discrimination task first, it is likely that participants will have been alerted to tones in the training phase. But this is a non-issue as long as no explicit information about the connections between the initial consonant and tone type will be provided.

included two kinds of items: critical items and extension items (Chan & Leung, 2014; 2018). Sound pairs for the critical and extension items (see details below) differed only in terms of the tone they carried (e.g. /p^hu:mR/ vs. /p^hu:mF/). 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. /p^hu:mR/ in the case above).

The critical items were concatenations of the following phonemes, resulting in 36 novel words (4x3x3):

- Onset: /p^h/, /t^h/, /l/ or /w/
- **Nucleus:** /u:/, /e:/ or /ɜ:/
- 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 connections 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 (2x6x3). They differed from those in the training phase *only in their onset consonants*:

- **Onset:** /k^h/, /j/
- Nucleus: /i:/, /u:/, /o:/, /ɛ:/, /ɔ:/ or /a:/
- 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 or other correlated rules; see the Discussion section for details), 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. The purpose of having 50% of the trials as fillers was to disguise the purpose of the test and discourage the use of explicit knowledge (Keating & Jegerski, 2015). To sum up, the pronunciation judgment tasks involved 144 trials, including 36 critical trials, 36 extension trials, and 72 filler trials.

The conscious state of participants' structural knowledge of the learning targets was also assessed via a source attribution task (Dienes, 2008). In each trial, after each pronunciation judgement, they were instructed to state the basis of their judgement as either "guess", "intuition", "recollection" or "rule/rules", which were defined as follows: "guess"—you were making a wild guess like flipping a coin; "intuition"—your judgement was based on a feeling that you cannot explain; "memory"—your judgement was based on a recollection; and "rule/rules": your judgement was based on one or more rules or partial rules that you can state. "Guess" and "intuition" attributions reflect unconscious structural knowledge, whilst "recollection" and "rule/rules" attributions reflect conscious structural knowledge (Dienes & Scott, 2005).

RESULTS

AX DISCRIMINATION TASK

We first examined participants' sensitivity to differences between sounds in the AB pairs (d') and potential bias towards certain responses (decision criterion) (c) based on Signal Detection Theory (Macmillan & Creelman, 2005), with 'Hit', 'Miss', 'False alarm' and 'Correct rejection' defined as shown in Table 2.

TABLE 2. Definitions of 'Hit', 'Miss', 'False alarm' and 'Correct rejection' for participants' responses in the AX discrimination task.

	Response: different	Response: same
Stimuli: AB pair	Hit	Miss
Stimuli: AA pair	False alarm	Correct rejection

Both d' and c were computed using the *psycho* package (Makowski, 2018) in R (R Core Team, 2012), and Figure 1 presents the average d' (left panel) and c (right panel) of each group. Most of the participants have a d' score over 3, suggesting that they were all highly sensitive to the differences between the sound pairs in the AB trials. One-way ANOVA reveals no significant difference in sensitivity among the four groups, $F(3, 76) = 0.154, p = 0.927$. On the other hand, the four groups appear to show a small variation in their response bias (c) and most participants' c score are all close to zero, revealing little overall response bias. One-way ANOVA reveals an overall significant difference in response bias, $F(3, 76) = 4.301, p = 0.007$. A post-hoc Tukey pair-wise multiple comparison (Table 3) reveals no significant difference for all comparisons except that English non-musicians had a significantly higher criterion for choosing "same" than Cantonese musicians. To sum up, these results are not surprising, as the two tones in the stimuli were meant to be easy for all participants to distinguish perceptually.

TABLE 3. Post-hoc Tukey pair-wise multiple comparisons for the response bias (c) of the four participant groups

Comparison	Estimated difference	Adjusted p
Cant non-musicians - Cant musicians	0.098	0.427
Eng musicians- Cant musicians	0.062	0.766
Eng non-musicians - Cant musicians	0.223	0.005*
Eng musicians - Cant non-musicians	0.036	0.945
Eng non-musicians - Cant non-musicians	0.125	0.215
Eng non-musicians - Eng musicians	-0.161	0.066

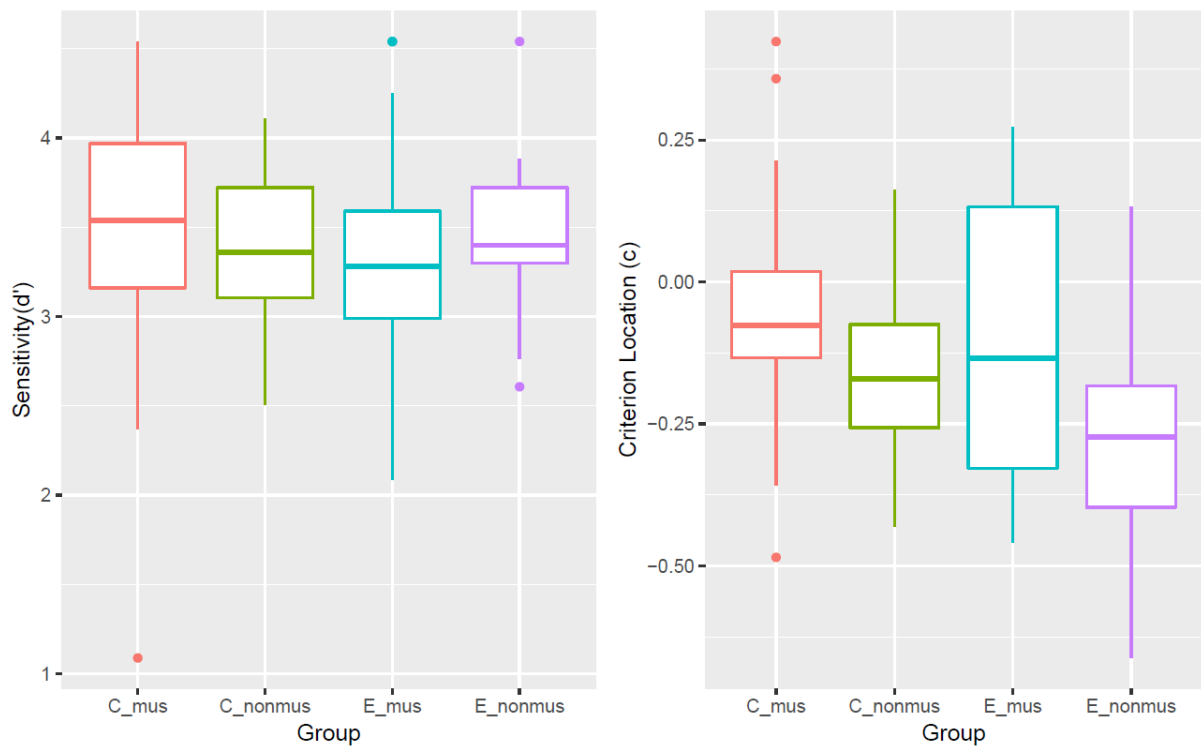


FIGURE 1. Sensitivity (d') (left) and response bias (c) (right) of the four participant groups in the AX discrimination task.

Since this task aimed to test participants' ability to distinguish the two tones in the learning targets, we now focus only on participants' responses to the AB pairs. The accuracy (no. of correct responses) and the reaction time data of the four participant groups were analysed. For each participant, reaction time values beyond ± 2.5 standard deviations from their

mean were treated as outliers and were excluded from analysis. The reaction time data were log-transformed for normalisation before any further statistical analyses. Figure 2 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).

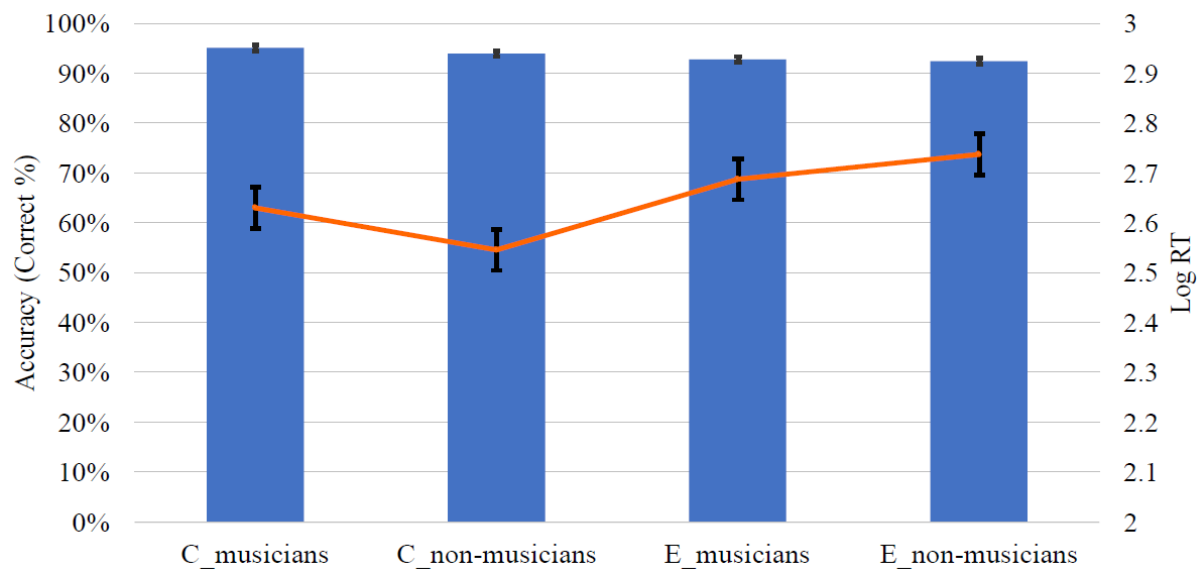


FIGURE 2. Accuracy and log-transformed reaction time of the four groups for the AB pairs in the AX discrimination task (C = Cantonese, E = English; error bar = 1 SE).

Mixed-effects models (LMMs) (R package *lme4* (Bates & Maechler, 2012) in R) were used to determine the effects of tonal experience and musical background (fixed effects) on participants' accuracy (generalised linear mixed-effects models, GLMMs) and reaction time (linear mixed-effects models). Participants and items (AB pairs) were included as random effects, with by-participant and by-item random slopes for the effect under investigation. For example, a full model for testing the effect of tonal experience was coded in R as “full_model <- glmer(Accuracy ~ Musical Background + Tonal Experience + (Tonal Experience |Participants) + (Tonal Experience|Item))”. Effects for each individual effect were tested by likelihood ratio

tests of a full model against a reduced model that excluded the effect to be tested⁶. Effects for interaction were tested by comparing models with and without the interaction (e.g. R code: *Musical background * Tonal experience* vs. *Musical background + Tonal experience*) with random intercepts for random effects. Results are presented in Table 4. 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 results suggest participants' accuracy was mostly not affected by their prior musical training and tonal experience, but tonal experience may have a relatively subtle facilitative effect for their reaction time, potentially at the automatized level. As expected, distinguishing a rising tone from a falling tone as in the learning targets was generally easy for our participants regardless of their tonal experience or musical background.

TABLE 4. Summary of the statistics of the mixed model comparisons for the factors of interest for the AX discrimination task (* = significant effect).

Factor	Result	
	Accuracy	Reaction time (log10)
Musical background	$\chi^2(1) = 0.562, p = 0.454^7$	$\chi^2(1) = 0.498, p = 0.480^8$
Tonal experience	$\chi^2(1) = 0.960, p = 0.327^9$	$\chi^2(1) = 21.06, p < 0.001^{*10}$
Musical background * tonal experience	$\chi^2(1) = 0.465, p = 0.495^{11}$	$\chi^2(1) = 99.08, p < 0.001^{*12}$

⁶ For example, to test the effect of musical background on accuracy, the R code for the full model is “full_model<-glmer(Accuracy ~ Musical Background + Tonal Experience+ (Musical Background|Participants) + (Musical Background|Item))”; the reduced model was coded as “reduced_model<- glmer(Accuracy ~ Tonal Experience + (Musical Background|Participants) + (Musical Background|Item))”

⁷ Full model: glmer(Accuracy ~ Musical Background + Tonal Experience+ (Musical Background | Participants) + (Musical Background | Item), family=binomial, control=glmerControl(optimizer = 'optimx', “optCtrl=list(method='nlnmb'”))”

⁸ Full model: (LogRT~ Musical Background + Tonal Experience + (Musical Background | Participants) + (Musical Background | Item), REML=FALSE)

⁹ Full model: glmer(Accuracy ~ Musical Background + Tonal Experience+ (Tonal Experience | Participants) + (Tonal Experience | Item), family=binomial, control=glmerControl(optimizer = 'optimx', “optCtrl=list(method='Nelder_Mead'”))”

¹⁰ Full model: (LogRT~ Musical Background + Tonal Experience + (Tonal Experience | Participants) + (Tonal Experience | Item), REML=FALSE)

¹¹ Full model: glmer(Accuracy ~ Musical Background * Tonal Experience+ (1 | Participants) + (1 | Item)). Any models with more complicated random slope structure failed to converge.

¹² Full model: lmer(LogRT ~ Musical Background * Tonal Experience+ (1 | Participants) + (1 | Item)). A random intercept model was used instead of a random slope model so that the results are comparable with the results from the model on accuracy.

PRONUNCIATION JUDGMENT TASK

Figure 3 shows the average accuracy of the four groups for the critical items and extension items. Impressionistically, for both critical items and extension items, English participants performed at around chance level (50%) but Cantonese participants performed considerably above chance level, regardless of their prior musical training.

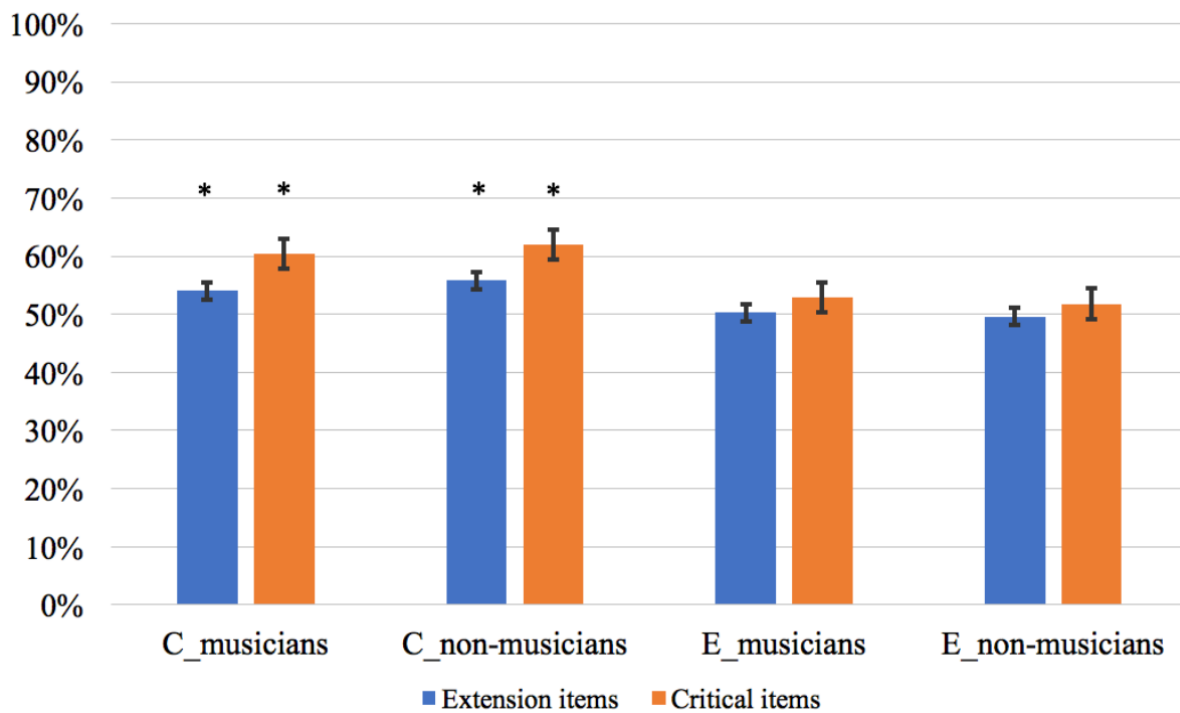


FIGURE 3. Accuracy of the four groups for the critical items and extension items in the pronunciation judgment task (C = Cantonese, E = English; error bar = 1 SE).

GLMMs were constructed to determine the effects of tonal experience, musical background and their interaction on participants' accuracy on the critical items and extension items with same procedure as above. Again, "Tonal experience" and "Musical background" were treated as fixed factors and "Item" and "Participant" as random factors in the models. Results are presented in Table 5, which reveal 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. No significant interaction between Musical background and Tonal experience was found.

TABLE 5. Summary of the statistics of the mixed model comparisons for the factors of interest for the pronunciation judgment task (* = significant effect).

Factor	Accuracy	
	Critical items	Extension items
Musical background ¹³	$\chi^2(1) = 0.001, p = 0.975$	$\chi^2(1) = 0.114, p = 0.915$
Tonal experience ¹⁴	$\chi^2(1) = 5.315, p = 0.0211^*$	$\chi^2(1) = 5.105, p = 0.0239^*$
Musical background * tonal experience ¹⁵	$\chi^2(2) = 0.482, p = 0.786$	$\chi^2(2) = 0.082, p = 0.960$

To determine whether individual groups performed significantly above chance level (i.e. showing a learning effect), further GLMMs were constructed with 'Participant Group' (4 levels) as the fixed factor and by-item and by-participant random intercepts¹⁶. 95% confidence intervals were generated for the prediction of each participant group based on GLMM using the "predict" function in R. Results are presented in Tables 6 and 7. 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 connections 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 did not perform significantly above chance level.

¹³ Full model: `glmer(Accuracy ~ Musical Background + Tonal Experience + (Musical Background | Participants) + (Musical Background | Item), family=binomial)`

¹⁴ Full model: `glmer(Accuracy ~ Musical Background + Tonal Experience + (Tonal Experience | Participants) + (Tonal Experience | Item), family=binomial)`

¹⁵ Full model: `glmer(Accuracy ~ Musical Background * Tonal Experience + (1 | Participants) + (1 | Item))`. Any models with more complicated random slope structure failed to converge.

¹⁶ R code: `glmer(Accuracy ~ Participant Group + (1 | Participants) + (1 | Item), family=binomial)`. There is no theoretical motivation for any random slope structures in the model.

TABLES 6 and 7. Accuracy (% correct) on and 95% confidence intervals (from upper bound to lower bound) of the critical items (top) and extension items (bottom) for the prediction of each participant group based on GLMM (chance level = 50%).

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

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

To determine whether each participant group performed significantly differently for words with different onset consonants and tones, another set of GLMMs were constructed, with ‘Item Type’ (2 levels: approximants vs. aspirated stops) as the fixed factor and by-item random intercept and by-participant random slope for the effect of Item Type. GLMMs were built separately for critical items and extension items. Results are presented in Table 8. For all the four participant groups, there is no significant difference in participants’ accuracy on approximant-initial words and aspirated-stop-initial words for either critical items or extension items, suggesting that in general our participants learnt the two target rules on tone-segment connections similarly well.

TABLE 8. Summary of the statistics of the mixed model comparisons on item types in the pronunciation judgment task (* = significant effect).

Participant Group	Overall accuracy	
	Critical items	Extension items
Cantonese musicians	$\chi^2(1) = 0.200, p = 0.655$	$\chi^2(1) = 2.804, p = 0.094$
Cantonese non-musicians	$\chi^2(1) = 3.124, p = 0.077$	$\chi^2(1) = 2.094, p = 0.148$
English musicians	$\chi^2(1) = 0.770, p = 0.380$	$\chi^2(1) = 0.923, p = 0.337$
English non-musicians	$\chi^2(1) = 2.872, p = 0.090$	$\chi^2(1) = 2.664, p = 0.103$

STRUCTURAL KNOWLEDGE

A further question is whether Cantonese musicians and non-musicians, who showed a learning effect of the target tone-segment connections, possessed relevant unconscious structural knowledge. Table 9 shows the number and percentage of responses of each source attribution (guess, intuition, memory, or rule(s)) for the critical items and extension items by the Cantonese musicians and non-musicians. For both critical items and extension items, over 60% of the source attribution responses were implicit attributions (i.e. guess or memory), with “guess” being the attribution with the highest frequencies. This suggests that most of the time the Cantonese participants were simply guessing or using their intuition for their judgments in the critical and extension trials. In general, there were more implicit attributions for extension items than critical items. This is in line with the fact that extension items involved new onset consonants which were not encountered in the training phase.

TABLE 9. Number and percentage of responses of each attribution for critical items and extension items (36 critical/extension items*40 participants = 1440 responses).

	Implicit attributions		Explicit attributions	
	Guess	Intuition	Memory	Rule(s)
Critical items	442 (30.7%)	438 (30.4%)	372 (25.8%)	188 (13.1%)
Extension items	511 (35.5%)	470 (32.6%)	281 (19.5%)	178 (12.4%)

To test whether Cantonese participants showed a learning effect of the target tone-segment connections even when they stated they were guessing or using their intuition, another set of GLMMs were built to determine participants’ accuracy when they stated “guess” or “intuition” (implicit attributions) and 95% confidence intervals were generated in the same way as reported above. Results are presented in Tables 10 and 11. 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 for both kinds of items even when they claimed to be guessing or

using their intuition. This reveals that Cantonese participants possessed implicit, abstract and potentially rule-like knowledge of the target tone-segment connections.

Tables 10 and 11. Accuracy (% correct) on and 95% confidence intervals (from upper bound to lower bound) of the critical items (top) and extension items (bottom) for the prediction of Cantonese participants (with data from their implicit attributions only) based on GLMM (chance level = 50%).

Group	Accuracy on critical items (%)	Upper Bound	Lower Bound
Cantonese musicians	58.56	63.78	53.14
Cantonese non-musicians	62.35	68.00	56.34

Group	Accuracy on extension items (%)	Upper Bound	Lower Bound
Cantonese musicians	55.65	59.94	51.28
Cantonese non-musicians	56.82	61.54	51.97

DISCUSSION

The present study sought to determine 1) whether tone-segment connections can be acquired through incidental exposure; and if so 2) whether the resultant knowledge may be abstract and implicit; and 3) the effects of prior musical training and tonal experience on the encoding of pitch patterns as abstract tone categories at the syllable level, as reflected in the learning of tone-segment connections. With regard to the goals (1) and (2), we found that Cantonese speakers, regardless of their musical background, developed knowledge of novel tone-segment connections only after brief exposure without explicit feedback. They performed similarly well on the two target rules for both critical items and extension items. With regard to the conscious state of the resultant knowledge, it was found that Cantonese speakers possessed implicit structural knowledge of the target tone-segment mappings as assessed by the source attribution task (Dienes, 2008). We conclude that it is possible for Cantonese learners (and potentially speakers of other tone languages as well) to acquire implicit knowledge of tone-segment connections through incidental exposure. This adds to an existing body of work demonstrating

that different linguistic patterns, especially phonological patterns, can be acquired by adult L2 learners (Chan & Leung, 2014, 2018; Dell et al., 2000; Grey et al., 2014; Leung & Williams, 2011, 2012, 2014; Rebuschat & Williams, 2012; Rogers et al., 2016; Warker & Dell, 2006; Williams & Kuribara, 2008).

The Cantonese speakers also appeared to have an abstract and potentially rule-like knowledge of the target tone-segment connections. In the present study, given the strict control on the frequency of phonemes in the nucleus and coda positions for the stimuli in the training phase and the testing phase, their above-chance performance on the critical items and the extension should have been based on their knowledge on the mapping between the onset consonants and the target tones. The Cantonese speakers displayed a learning effect for novel words whose onset consonants were encountered in the training phase (critical items), suggesting that they possessed knowledge of the target tone-segment connections which should be at a sufficient level of abstraction to facilitate transfer to completely new lexical items rather than relying solely on their memory of the training items. Moreover, they also showed a learning effect for novel words with onset consonants not encountered in the training phase (extension items), suggesting that their knowledge is potentially rule-like. While they may have acquired rules described in terms of the phonological classes of the onset consonant and the tone types (i.e. words beginning with an aspirated stop always carry a rising tone, and words beginning with an approximant always carry a falling tone), they might also have learnt rules about the mappings of tones with other salient by-products of the target rules (correlated rules) (e.g. words beginning with an aspiration or a “puff of air” always carry a rising tone, or words beginning with a consonant with vowel-like quality (a property of approximants) always carry a falling tone). Although the design of the present study did not allow us to determine whether their knowledge was based on the entire phonological classes or the associated features of these phonological classes, our findings provide evidence that their knowledge of the target tone-

segment connections was abstract and potentially rule-based knowledge instead of merely based on memory of the training items. Still, it should be noted that Cantonese speakers' performance on extension items was worse than that on critical items. This is in line with previous studies which used the transfer tests (e.g. Altmann et al., 1995): judgment performance was worse on test items with different surface features (e.g. different letter set) or in different modalities (e.g. from visual to auditory). Specifically in our study, one main difference between the onset consonants in the training/critical items and those in the testing items lies in the feature [anterior]: onset /p^h/, /t^h/, /l/ and /w/ in the training/critical items were [+anterior] whereas onset /k^h/ and /j/ were [-anterior]. Participants might thus be less able to apply the knowledge they had acquired in the training to the extension items. This interpretation is consistent with the fact that all four participant groups apparently had lower accuracy on the extension items than on the critical items.

As for goal (3), our study demonstrates that incidental learning of tone-segment connections does not appear to be completely unconstrained. The English speakers in our study, be they musicians or non-musicians, did not show any learning effect of the target tone-segment connections. This does not mean that tone-segment connections are not learnable for them, as we cannot rule out the possibility that the target tone-segment connections could also be learnable by English speakers if more input was given (e.g. more trials in the training phase). Still, with the same amount of input and similar experiment duration, our results show that Cantonese speakers and English speakers exhibited differential success in learning the target tone-segment connections via incidental exposure. What could have caused such a difference in their performance? Our learning targets involved the mappings between onset consonants (aspirated stops vs. approximants) and lexical tones (rising or falling). At the segmental level, both aspirated stops and approximants are phonological natural classes in English and Cantonese; both Cantonese and English speakers in our study should share a similar starting

point with regard to the segmental parts of the target rules. Besides, both English and Cantonese speakers performed similarly well in distinguishing between the rising tone and the falling tone in the learning targets as revealed in the AX discrimination task, implying that the observed differential performance was unlikely to be due to difficulty in perceiving tones at the phonetic-acoustic level. Then it is likely that the observed differential performance was due to how the tones in the learning targets were processed, which can be explained by the Perceptual Assimilation Model (Best, 1994). Specifically, our Cantonese speakers, who also spoke Mandarin as an L2, may have perceptually assimilated the rising and falling tones in the learning targets to the Cantonese and/or Mandarin counterparts. As such, they could encode novel pitch patterns as abstract tone categories at the word level, which was the prerequisite of learning novel tone-segment connections incidentally. In contrast, English learners may have perceptually assimilated the rising and falling tones in the learning targets to intonational rise and fall in the English tonal system (So & Best, 2010). Although contrastive intonational patterns may also operate on monosyllabic words in English and other non-tonal languages, intonation usually operates at the utterance level and is not used for contrasting word meaning. Therefore, English learners might not be able to encode novel pitch patterns as abstract tone categories at the syllable level and thus did not show a learning effect of the target tone-segment connections.

These results have implications for L2 lexical tone learning in general. L2 lexical tones are often reported to be difficult, especially for learners whose L1 is a non-tonal language, but the nature of the long-term difficulty involved remains far from clear (Pelzl et al., 2018). Most previous studies have focused on the perception of L2 lexical tones at the phonetic-acoustic level (e.g. Francis et al., 2008; Hallé et al., 2004; Wayland & Guion, 2004). However, it is not until the past decade that the distinction between phonetic perception/processing of sounds and phonological and lexical perception/processing of tones has been highlighted in the L2 tone

learning literature (e.g. Malins & Joanisse, 2010, 2012; Pelzl et al., 2018; Wiener et al., 2018). For instance, Pelzl et al. (2018) found a discontinuity of success in L2 tone learning: they found that L1 English advanced learners of Mandarin performed at a near-native level in identifying tones on isolated words, but they struggled in a lexical decision task and a semantic judgment task which hinged on lexical tone processing. A possible reason lies in English learners' inability to "repurpose" pitch patterns from intonational cues to lexical cues, but they must first form abstract tone categories at the word level, which can be a separate challenge for English and other non-tonal language speakers, before they can use them as lexical cues in lexical tasks. However, previous studies on the 'lexical learning' of tones conflate lexical tone processing at the phonological level (abstract and categorisation) with the lexical level (processing of word meaning), by using lexical tasks to infer participants' ability to establish novel phonological categories. As far as we aware, this is the first study that teases apart the phonological level from the phonetic and lexical levels in tone processing and demonstrated that forming abstract tone categories at the word level may be difficult for English speakers, which is a prerequisite of using abstract tone categories as lexical cues for contrasting word meaning. Specifically, in the present study, English speakers achieved native-like accuracy in distinguishing the rising tone from the falling tone in the learning targets but they nevertheless failed to learn the target tone-segment connections. Such "discontinuity" is potentially due to their processing bias in treating pitch patterns as intonational cues. 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 both the overlap and separation of the musical and linguistic domains. (see section 1.3 for a detailed

discussion). However, a crucial question about the relationship between music and speech remain unexplored: that is, whether prior musical training may facilitate the formation of novel lexical tone categories. The present study shows that prior musical training did not have any significant effect on the learning of tone-segment connections by either Cantonese or English speakers, suggesting that prior musical training does not facilitate the encoding of pitch patterns as abstract tone categories at the syllable level. This is in line with a previous finding that, while musicians who speak a non-tonal language were relatively more accurate in pitch contour identification than non-musicians, their pitch contour abstraction and categorization ability was comparable to that of non-musicians (Wayland et al., 2010), suggesting that musicians do not have an advantage in forming abstract categories based on pitch patterns despite better ability to perceive tones at the phonetic-acoustic level. Our results highlight an area of the separation between the musical domain and the linguistic domain.

At a broader level, our findings reveal that the incidental learning mechanism, despite being domain-general, does not seem to involve unconstrained associative learning but may interact with learners' prior linguistic experience and processing biases. Specifically, the incidental learning of tone-segment connections appears to depend crucially on how pitch variation is used in the tone system(s) of the language(s) the learners speak, which shapes their processing bias on pitch patterns. The present study adds to a relatively small body of work that investigates the limits of and constraints involved in incidental/implicit learning (e.g. Leung & Williams, 2011; 2012; 2014).

In the present study, it was our intention to 1) use simple tonal contrast (i.e. rising vs. falling) in our stimuli which were easy for all participant groups to discriminate at the phonetic-acoustic level; 2) have strict control on the frequency of all the different phonemes in the stimuli; and 3) make the tasks entirely auditory without involving meaning in the stimuli. With careful experimental control on the stimuli based on the learning targets, participants were exposed to

systematic data and the results provide strong evidence that the differential success by Cantonese and English speakers in the incidental learning of tone-segment connections can only be attributed to how tones were processed at the *phonological* level. Due to the quest for optimum experimental control, one might imagine that since the tone-segment connections involved in our study were artificial and simple when compared with tone-segment constraints found in some natural languages such as Thai (Sladen, 2009), the findings seem to have little apparent relevance to L2 speech learning in naturalistic settings. However, as in many psychology/psycholinguistic experiments, a plausible assumption is that participants brought to the laboratory the same processing resources that they use in everyday life, and thus their performance must reveal something about how their minds work. In terms of language learning, it has been found that success in an artificial language learning experiment correlates positively with indices of L2 learning (Ettliger, Morgan-Short, Faretta-Stutenberg, & Wong, 2016). Therefore, it is reasonable to believe that our findings are also relevant to L2 tone learning in naturalistic settings.

CONCLUSION

The present study investigated the incidental learning of tone-segment connections by Cantonese and English speakers. Despite similar performance on distinguishing the rising tone from the falling tone perceptually, Cantonese speakers showed an abstract and potentially rule-like knowledge of the target tone-segment connections but English speakers did not, attributable to the possibility that English learners failed to form tone categories at the syllable level. Also, prior musical training did not seem to facilitate the formation of lexical tone categories, providing evidence for the separation between music and speech.

REFERENCES

- Andringa, S., & Rebuschat, P. (Eds.) (2015). New directions in implicit and explicit language learning [special issue]. *Studies in Second Language Acquisition*, 37.
- 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.
- Bates, D.M., & Maechler, M. (2009). *lme4: Linear Mixed-Effects Models Using S4 Classes*, R Package Version 0.999375-32.
- Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Walter de Gruyter.
- Best, C. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In C. Goodman & H. Nusbaum (Eds), *The development of speech perception*, 167-224. Cambridge: The MIT Press.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and cognition*, 77(1), 1-10.
- Bowles, A. R., Chang, C. B. & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, 66(4), 774–808.
- Bradley, E. D. (2012). *Crosslinguistic Perception of Pitch in Language and Music*. Ph.D. dissertation, University of Delaware, Newark, DE.
- Chan, R. & Leung, J. (2014). Implicit learning of L2 word stress regularities. *Second Language Research*, 30(4), 463-484.
- 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.
- Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America*, 128(1), 456–465.
- Chang, C. B., & Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *Journal of the Acoustical Society of America*, 136(6), 3703–3716.

- Creel, S. C., Weng, M., Fu, G., Heyman, G. D., & Lee, K. (2018). Speaking a tone language enhances musical pitch perception in 3–5-year-olds. *Developmental science*, *21*(1), e12503.
- Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1355.
- Deutsch, D., Dooley, K., Henthorn, T., & Head, B. (2009). Absolute pitch among students in an American music conservatory: Association with tone language fluency. *The Journal of the Acoustical Society of America*, *125*(4), 2398-2403.
- Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception: An Interdisciplinary Journal*, *21*(3), 339-356.
- Dienes, Z. (2008). Subjective measures of unconscious knowledge. *Progress in Brain Research*, *168*, 49–64.
- Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*, 338–351.
- Ettliger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. (2016). The relationship between artificial and second language learning. *Cognitive Science*, *40*, 822–847.
- 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.
- Gandour, J., Dziedzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., ... & Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and language*, *84*(3), 318-336.
- Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition*, *38*(2), 177-215.
- Gomez, R. L. & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–35.

- Gottfried, T. L. (2007). Music and language learning: Effect of musical training on learning L2 speech contrasts. In O. S. Bohn & M. Munro (Eds.), *Language Experience in Second Language Speech Learning*, (pp. 221–237). Philadelphia: John Benjamins.
- Graham, C. R., & Williams, J. N. (2016). Implicit learning of Latin stress regularities. *Studies in Second Language Acquisition*, 1-27.
- Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, 36, 1–34.
- 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.
- Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40, 269–279.
- Hulstijn, J. H. (2011). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2632-2640). Chichester: Wiley-Blackwell.
- Hulstijn, J. H., & Ellis, R. (Eds.) (2005). Implicit and explicit second-language learning [special issue]. *Studies in Second Language Acquisition*, 27.
- Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia*, 48(9), 2630-2639.
- 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.
- Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25(1), 161-168.
- Leather, J. (1987). F0 pattern inference in the perceptual acquisition of second language tone. In J. Leather & A. James (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 59–80). Providence: Foris.

- Lee, C. Y., & Hung, T. H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*(5), 3235-3248.
- Lee, C. Y., Lee, Y. F., & Shr, C. L. (2011). Perception of musical and lexical tones by Taiwanese-speaking musicians. *The Journal of the Acoustical Society of America*, *130*(1), 526-535.
- Lee, C. Y., Tao, L., & Bond, Z. S. (2009). Speaker variability and context in the identification of fragmented Mandarin tones by native and non-native listeners. *Journal of Phonetics*, *37*(1), 1-15.
- Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics*, *53*(2), 157-165.
- Lee, W. S., & Zee, E. (2009). Hakka Chinese. *Journal of the International Phonetic Association*, *39*(1), 107-111.
- Leung, J. & Williams, J. (2011) The implicit learning of mappings between forms and contextually-derived meanings. *Studies in Second Language Acquisition*, *33*(1): 33-55.
- Leung, J. & Williams, J. (2012) Constraints on implicit learning of grammatical form-meaning connections. *Language Learning*, *62*(2): 634-662.
- Leung, J. & Williams, J. (2014) Cross-linguistic differences in implicit language learning. *Studies in Second Language Acquisition* *36*(4): 733-755.
- Macmillan, Neil A., & Creelman, C. Douglas (2005). *Detection Theory: A User's Guide*. Lawrence Erlbaum: Mahwah.
- Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407-420.
- Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, *50*(8), 2032-2043.

- 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.
- Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain*, 133(9), 2635-2642.
- Norman, J. (1988). *Chinese*. Cambridge: CUP.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, 2, 142.
- Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2018). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition*, 1-28.
- Perrachione, T. K., Fedorenko, E. G., Vinke, L., Gibson, E., & Dilley, L. C. (2013). Evidence for shared cognitive processing of pitch in music and language. *PLOS ONE*, 8(8), e73372.
- Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America*, 130(1), 461-472.
- R Core Team. (2013). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Version 3.0.0., <<http://www.R-project.org>>.
- Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford: Oxford University Press.
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, 63(3), 595-626.
- Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 33(4), 829-856.
- Repp, B. H., & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception: a cross-linguistic study. *Journal of Phonetics*, 18, 481-495.
- Rogers, J., Revesz, A., & Rebuschat, P. (2016). Implicit and explicit knowledge of inflectional morphology. *Applied Psycholinguistics*, 37(4), 781-812.
- Sagart, L. (1999). The origin of Chinese tones. In *Proceedings of the Symposium/Cross-Linguistic Studies of Tonal Phenomena/Tonogenesis, Typology and Related Topics*. (pp.

- 91-104). Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Psycholinguistics*, *11*, 129–158.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.
- Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, . . . I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721–737). Singapore: National University of Singapore, Centre for Language Studies.
- Scott, R. & Dienes, Z. (2008) The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *34*, 1264-88.
- Sebastián-Gallés, N., & Díaz, B. (2012). First and second language speech perception: Graded learning. *Language Learning*, *62*(s2), 131–147.
- Sladen, G. (2009). *Central Thai phonology*. Retrieved from <http://www.thailanguage.com/resources/slayden-thai-phonology.pdf>
- 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.
- Stagray, J. R., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and nontone language. *Journal of Chinese Linguistics*, *21*, 143–163.
- Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics*, *25*(3), 449-466.
- Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin Tones by Chinese and American listeners. *Brain and Language*, *78*(3), 332–348.
- 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.

- Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 387.
- 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., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, 38(4), 654-662.
- Wiener, S., Ito, K., & Speer, S. R. (2018). Early L2 Spoken Word Recognition Combines Input-Based and Knowledge-Based Processing. *Language and speech*, DOI: 0023830918761762.
- Wells, J. (2006). *English intonation: An introduction*. Cambridge: Cambridge University Press.
- Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The New Handbook of Second Language Acquisition*, (pp. 319-353). Emerald Group Publishing Limited.
- Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua*, 118, 522-553.
- Wong, P. C. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain research bulletin*, 59(2), 83-95.
- Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28(4), 565-585.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience*, 10(4), 420.
- Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, 120(2), 1063-1074.
- Zhang, L. (2011). Meiguo liuxuesheng Hanyu shengdiaode yinwei he shengxue xinxi jiaogong. *Shijie Hanyu Jiaoxue Chinese Teaching in the World*, 25, 268-275.