Brain & Language 229 (2022) 105106

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

The tone atlas of perceptual discriminability and perceptual distance: Four tone languages and five language groups

Liquan Liu [a,b,c,d,*], Regine Lai [e], Leher Singh [f], Marina Kalashnikova [b,g,h], Patrick C.M. Wong [e,i], Benjawan Kasisopa [b], Ao Chen [j], Chutamanee Onsuwan [k], Denis Burnham [b,*]

[a] School of Psychology, Western Sydney University, Australia
[b] The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia
[c] Center for Multilingualism in Society Across the Lifespan, University of Oslo, Norway
[d] Centre of Excellence for the Dynamics of Language, Australian Research Council, Australia
[e] Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong
[f] Department of Psychology, National University of Singapore, Singapore
[g] Basque Center on Cognition, Brain and Language, Spain
[h] Ikerbasque, Basque Foundation for Science, Spain
[i] Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong
[j] School of Communication Sciences, Beijing Language and Culture University, China
[k] Department of Linguistics, Faculty of Liberal Arts and Center of Excellence in Intelligent Informatics, Speech and Language Technology, and Service Innovation (CILS), Thammasat University, Thailand

ARTICLE INFO

Keywords:
Tone
Cross-linguistic perception
Tone system
Contrast type
Perceptual asymmetry
Cue-weighting
Multi-dimensional analysis

ABSTRACT

Some prior investigations suggest that tone perception is *flexible*, reasonably independent of native phonology, whereas others suggest it is *constrained* by native phonology. We address this issue in a systematic and comprehensive investigation of adult tone perception. Sampling from diverse tone and non-tone speaking communities, we tested discrimination of the three major tone systems (Cantonese, Thai, Mandarin) that dominate the tone perception literature, in relation to native language and language experience as well as stimulus variation (tone properties, presentation order, pitch cues) using linear mixed effect modelling and multidimensional scaling. There was an overall discrimination advantage for tone language speakers and for native tones. However, language- and tone-specific effects, and presentation order effects also emerged. Thus, over and above native phonology, stimulus variation exerts a powerful influence on tone discrimination. This study provides a tone atlas, a reference guide to inform empirical studies of tone sensitivity, both retrospectively and prospectively.

1. Introduction

Around 60–70% of the world's languages are tone (or pitch-accent) languages – they use pitch to differentiate word meanings (Yip, 2002). Over half of the world's population speaks a tone language (Fromkin, 1978). Despite this ubiquity, the determinants of tone perception are not well understood. Some suggest tone perception is flexible, reasonably independent of native phonology, whereas others suggest it is constrained by native phonology or perceptual asymmetries. This study aims to investigate the relative roles of the perceiver, the stimulus and the context in lexical tone perception. To this end, the study involves a comprehensive investigation of how different tones in different tone languages are perceived by listeners from different language backgrounds in different task contexts.

In terms of the classification of lexical tones, unlike other phonetic units (consonants, vowels), which are often defined by articulatory gestures, tones are defined mostly by their pitch characteristics (Burnham, Attina, Xu, & Kasisopa, 2011). Fundamental frequency (f_0), the acoustic basis of pitch perception, is a widely recognized index of tone differences (Barry & Blamey, 2004), even though tones are also differentiated, to a lesser extent, on other dimensions, such as duration, amplitude, and voice register (Feng, Wu, & Nissenbaum, 2020). Given the predominance of pitch in the perception of tone, Chao (1930) introduced a system to describe individual tones, whereby each tone is

* Corresponding authors.
E-mail addresses: l.liu@westernsydney.edu.au (L. Liu), Burnham@westernsydney.edu.au (D. Burnham).

https://doi.org/10.1016/j.bandl.2022.105106
Received 15 March 2021; Received in revised form 2 March 2022; Accepted 8 March 2022
Available online 4 April 2022

described on a normalized pitch scale from 1 (low) to 5 (high), one value at the start, another at the end, and a third medial value when there is an inflexion point within the tone. As this index is normalized, it is independent of the language and speaker and is used here to describe and compare tones within and between languages.

The core issue in studying perceptual discriminability and perceptual distance lies in the understanding of the determinants of tone discrimination. Listeners' tone perception has been shown to be influenced by multiple factors. Here, we discuss the effects of the perceiver (tone language experiences); the stimulus (including tone systems, contrast type) and the stimulus itself (tone properties and cues); and the task context (order of stimulus presentation) on tone perception.

Like the perception of other phonetic units, tone perception is shaped by linguistic experience from the beginning of life (Fikkert, Liu, & Ota, 2020; Werker, 2018). Unlike consonants and vowels for which a clear developmental trajectory has been reported involving perceptual narrowing, i.e., from flexible (cross-language, phonetic) to constrained (language-specific, phonological) perception, the picture is mixed for tones. Some studies have demonstrated perceptual narrowing, indexed by decreased tonal sensitivity over age in non-native tone learners (Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008; Quam & Swingley, 2010). Others have reported facilitation, an age-based *increased* sensitivity by both tone and non-tone language learning infants (Chen & Kager, 2016; Chen, Stevens, & Kager, 2017; Ramachers, Brouwer, & Fikkert, 2018; Singh et al. 2018; Tsao, 2017). Yet other studies report results that attest to the flexibility of tone perception – a renewed sensitivity to tone by non-tone language learning infants in their second year (Götz, Yeung, Krasotkina, Schwarzer, & Höhle, 2018; Liu & Kager, 2014, 2017). Burnham and Singh (2018) point out that the mixed results are the product of studies using different tone pairs and different tone languages and could well be due to the psychophysics of tone salience. Accordingly, more acoustically salient tones may be more resilient to listeners across language backgrounds in the course of perceptual narrowing. However, the measure of inherent tonal salience has been difficult and is an issue of controversy, exacerbated by the absence of tone perception metrics solidly grounded on articulatory gestures.

The effect of language background is evident in studies of adult listeners' tone perception, in which tone language speaking adults show distinct and enhanced tone discrimination compared with their non-tone language-speaking peers (Burnham & Francis, 1997; Kaan, Wayland, Bao, & Barkley, 2007; Malins & Joanisse, 2012; also see Maggu, Zong, Law, & Wong, 2018 and Tong, Lee, Lee, & Burnham, 2015 for the effect of monolingual versus bilingual tone and non-tone languages). In a study measuring Thai tone perception by speakers of Thai, Cantonese, Mandarin (tone languages), Swedish (pitch-accent) and English, better perception was evident in native speakers of a tone/pitch-accent language over native non-tone (English) speakers (Burnham et al., 2015).

Direct comparisons of tone language speakers' native versus non-native tone perception are rare, but indirect evidence suggests differences in the perception of native and non-native tones. As early as 4 months after birth, Cantonese- and Mandarin-learning infants display language-specific perception of Cantonese tones (Yeung, Chen, & Werker, 2013). Such an effect also appears to carry over into adulthood: While Thai speakers have no difficulty identifying Thai tones, Mandarin and Cantonese speakers' Thai tone perception appears to be modulated by their own tonal inventory (Reid et al., 2015). It has been argued that listeners perform better when non-native tones can be assimilated to their native tone categories (Chen, Best, & Antoniou, 2020), thus adhering to assimilation accounts of native versus non-native speech perception (e.g., Perceptual Assimilation Model, Best, 1995). These combined results strongly suggest, but do not directly test the hypothesis, that tone language speakers' perception of native tones should be superior to their perception versus non-native tones In this paper, we will directly test the hypothesis that tone language speakers' perception of native tones will be better than their perception of non-native tones.

Tone language speakers' processing of tones involves not only the bilateral frontoparietal brain regions used for acoustic analysis and abstraction (Gandour, Dzemidzic, et al., 2003; Gandour, Wong, et al., 2003; Gandour et al., 2004; Li et al., 2010), but also frontotemporal regions used for processing phonological and semantic information (Kwok, Dan, Yakpo, Matthews, & Tan, 2016; Kwok et al., 2017; Liang & Du, 2018; Wong et al., 2004). In addition, tone processing in listeners from tone language backgrounds becomes more left-lateralised when language comprehension tasks evoke more semantic processing (Gandour, Dzemidzic, et al., 2003; Gandour, Wong, et al., 2003; Gandour, Wong, & Hutchins, 1998; Gandour et al., 2004). Moreover, Chen, Peter, et al. (2018) found that while tone (Mandarin Chinese) and non-tone (Dutch) language adults show similar right-lateralised mismatch negativity (MMN) for 3-note musical melodies based on Mandarin tones, for the lexical tones on which the f_0 of these 3-note melodies were based, Chinese adults showed a later MMN peak to lexical tone oddballs than did non-tone Dutch adults. These results imply that, consistent with findings that tone language speakers process tones categorically (Feng, Gan, Wang, Wong, & Chandrasekaran, 2018; Gandour & Krishnan, 2016; Peng et al., 2010; Xi, Zhang, Shu, Zhang, & Li, 2010), a larger f_0 difference is necessary for Chinese than for Dutch adults to detect the lexical tone change. Moreover, Chen and colleagues (Chen, et al., 2018; Chen, Liu, & Kager, 2016; Liu, Chen, & Kager, 2020) found significant correlations between lexical tone and music pitch for non-tone, but not tone language speakers. It is known that adult-like right-hemispheric lateralisation to piano tones develops early at 2 months after birth (He, Hotson, & Trainor, 2007), and therefore, given this and the above findings, it appears that at the neurophysiological level non-tone language adults perceive lexical tones in a similar fashion as they do non-lexical (musical) pitch variations, whereas adult tone language speakers shift the manner of processing of lexical (but not musical) tones sometime after early infancy, and this facilitates the categorisation of f_0 modulations that signal phonological distinctions.

At the behavioural level, both the absence and presence of tone-relevant experience affects non-tone language speakers' lexical tone perception. On the one hand, the *absence* of tone-relevant experience *limits* tone discrimination (Burnham & Mattock, 2007). Burnham et al. (2015) showed that lexical tones presented in a speech context were less well discriminated by non-tone (English) language listeners than by tone or pitch-accent listeners, but once the very same f_0 pitch contours were presented as in a non-linguistic context, as hums or violin glides, differences in the perceptual performance between tone- and non-tone language listeners were diminished. On the other hand, the *presence* of tone-relevant experience shows that tone discrimination is *flexible*; adult non-tone language listeners' tone perception is improved by both tone training (Antoniou & Wong, 2015; 2016; Chandrasekaran, Sampath, & Wong, 2010; Francis, Ciocca, Ma, & Fenn, 2008; Ingvalson, Barr, & Wong, 2013) and musical training (Burnham, Brooker, & Reid, 2015; Cooper & Wang, 2012; Wong & Perrachione, 2007). Thus, while there is developmental divergence in lexical tone perception as a function of language environment, experience with a non-tone versus a tone language does not result in complete tone insensitivity in infants. In adults, non-tone language listeners' tone discrimination is a function of the acoustic continuum of f_0, and can be improved as a function of tone-relevant training and experience.

With respect to the tone language system, it is unclear whether certain systems are easier to perceive than others. From a phonological density perspective, it is reasonable to predict that tone systems with less density (i.e., fewer tones) would be easier to perceive. There is an additional factor that is related to the tone language system: the type and mix of tone types in a tone language. Lexical tones have been classed as static or dynamic (Abramson, 1978). Static tones (S) are those with a relatively level contour (e.g., Cantonese 55) or a mild slope (e.g., Thai 45), differing from one another mainly in pitch height; dynamic tones (D) involve distinct pitch contours (changes in pitch height over time), such as dynamic rising, (Dr, e.g., Cantonese 25), dynamic falling (Df, e.g., Mandarin 51), or more complex contours (e.g., rise-fall, Thai 241). In terms of contrast type, there is evidence from studies examining multiple tone contrasts that listeners' tone discrimination is related to the type of tone contrast they hear. For instance, in Huang and Johnson (2010), discrimination accuracy of the Mandarin rising-falling 35–51 tone contrast was the best, suggesting that dynamic rising versus dynamic falling tones may be the easiest to discriminate. Similarly, in a comparison across tone contrast types, Burnham and colleagues found that Dr-Df tone pairs are more easily discriminated than S-S or S-D tone pairs (Burnham, Kirkwood, Luksaneeyanawin, & Pansottee, 1992; Burnham, Kasisopa, Reid, Luksaneeyanawin, Lacerda, Attina, Xu Rattanasone, Schwarz, & Webster, 2015). Moreover, it appears that the static versus dynamic distinction has psychological reality, for it has been reported that listeners' native static tone experience facilitates their perception of static tones in another tone language (Chiao, Kabak, & Braun, 2011; Qin & Mok, 2011). More comprehensive empirical studies investigating more contrast types across languages are needed.

Another factor that can influence perception is the order in which stimuli are presented. "Perceptual asymmetry" occurs when the order of presentation leads to asymmetrical perceptual patterns. This phenomenon has been reported for segmental features, especially vowels (Polka, Ruan, & Masapollo, 2018; Zhao, Masapollo, Polka, Ménard, & Kuhl, 2019). With respect to lexical tones, Chen, Liu and Kager (2015) have shown that Mandarin tone sandhi (e.g., articulation of dipping-dipping 214–214 tone sequences as rising-dipping, 35–214) may rest upon a perceptual effect – directionally-specific acoustic masking. In line with this, it has been found that Mandarin rising-to-dipping, 35–214 and rising-to-falling, 35–51 tone pairs elicit larger MMN responses than the same tones in the opposite order (214–35 and 51–35), among native but not non-native Mandarin speakers, suggesting a perceptual constraint related to stimulus order on native listeners' phonological representations (Politzer-Ahles, Schluter, Wu, & Almeida, 2016). Similarly, Wayland and Chen (2018) found changes from Mandarin flat (55) to other tones quite challenging, and from Mandarin falling (51) to other tones easier than the reverse order (other tones to 51) for both tone (Mandarin) and non-tone (English) speakers. However, a follow-up study reported the opposite pattern for Mandarin flat tones, and a change from Mandarin dipping to other tones is more difficult than the reverse direction (Wayland, Chen, Zhou, & Hong, 2019). Moreover, processing load (Liu, Ong, Tuninetti, & Escudero, 2018) and acoustic cues such as spectral dynamicity (Masapollo, Zhao, Franklin, & Morgan, 2019) or breathiness (Yang & Sundara, 2019) may all play a role in whether perceptual asymmetry occurs in tone discrimination. Theoretical models have been proposed to account for the constraints imposed on tone perception by the asymmetry of the order effects, including the under-specification hypothesis (e.g., Politzer-Ahles et al., 2016), prototypicality theory (Best, 1994), natural referent vowel framework (Polka & Bohn, 2003) and psychophysics of tone salience (Burnham & Singh, 2018). Further study is required to investigate the relative merit of these theories and potential constraints through empirical research.

Finally, with respect to tone properties, pitch height and pitch direction have consistently been argued to be the two dominant cues in tone perception (Gandour, 1983; Gandour & Harshman, 1978; Lin, 1987; Vance, 1977) and recognition (Howie, 1976; Xu, 1997). While this is the case for both tone and non-tone language listeners, functional magnetic resonance imaging (fMRI) (Feng et al., 2018) and multidimensional scaling (MDS) studies (Chandrasekaran et al., 2010; Francis et al., 2008) have shown that tone language speakers attend more to directional cues and non-tone language speakers more to height cues, indicating a native tonological/phonological constraint (Chandrasekaran et al., 2010; Francis et al., 2008).

Here, in order to extract general principles and specific drivers of tone perception and to investigate flexibility and constraint in tone processing, listeners from four different tone language backgrounds (Chinese Mandarin, Singaporean Mandarin, Hong Kong Cantonese, and Bangkok Thai) and one non-tone language (Australian English) were tested on tone discrimination in four different tone systems (Chinese Mandarin, Singaporean Mandarin, Cantonese, and Thai), all of which have both static and dynamic tones but differ in the number of tones. We expected superior discrimination of lexical tones by tone- over non-tone language speakers and, among tone language speakers, of native over non-native tones. We also expected acoustic properties of tone would modulate listeners' perceptual ability for tones across language systems and tone types (Choi, Tong, & Singh, 2017). As mixed findings have been reported regarding perceptual asymmetry, we keep the prediction open for whether and in what direction tonal asymmetries may occur. Finally, we focused on the two primary dimensions of tone perception, pitch height and pitch direction, exploring the role of 23 acoustic measurements/cues in these two dimensions for tone perception (see Appendix III) and predicted that listeners would use the cues and/or pitch-related knowledge that are relevant in their native language to non-native tone perception. To increase the validity and generalisability of the study, two discrimination tasks, AX and AXB, were used in two separate groups of participants. The results for the AX task are presented in the body of the paper, with the methods of the AXB task reported in Supplementary material along with any AXB results that showed meaningful divergent patterns from the AX task results.

2. Methods

2.1. Participants

One hundred and twenty adults from five language backgrounds participated in an AX discrimination task ($N = 24$ per language background): Cantonese speakers ($M = 21.00$ years, $SD = 1.64$) tested at the Chinese University of Hong Kong, Thai speakers ($M = 27.33$ years, $SD = 5.41$) tested at the Thammasat University, Mandarin Chinese speakers ($M = 21.25$ years, $SD = 2.44$) tested at the Beijing Language and Culture University, Singaporean Mandarin speakers ($M = 21.70$ years, $SD = 2.18$) tested at the National University of Singapore, and Australian English speakers ($M = 28.38$ years, $SD = 10.43$) tested at the Western Sydney University (see Appendix I for details including a separate group of 120 participants for AXB). Tone language speaker groups included those who knew other languages including tone/pitch-accent languages, but all used their native tone language in their daily lives. Australian English speakers had no prior tone language experience.

2.2. Stimuli

All tones from the four tone systems, Cantonese (6 tones, 15 pairwise contrasts), Thai (5 tones, 10 contrasts), Chinese Mandarin (4 tones, 6 contrasts) and Singaporean Mandarin (4 tones, 6 contrasts), were used as stimuli, resulting in 37 within-language tone contrasts (Appendix II). Tones were categorized into three types: static (S) tones, dynamic rising (Dr) tones, and dynamic falling (Df) tones, resulting in five contrast types: S-S, S-Df, S-Dr, Df-Dr, Dr-Dr. To increase generalisability, each tone was presented on six different CV syllables (two initials (/p, pʰ/) × three vowels (/i, a, u/)). All stimuli ($N = 90$) words or phonologically legal non-words in their respective languages, were spoken by a native speaker of each language (all young females: Cantonese, 24 years; Chinese Mandarin, 30 years; Singaporean Mandarin, 28 years; Thai, 27 years). The speaker produced six to eight tokens of each tone word. Two most representative tokens (not the first or the last to avoid 'list' intonation, or productions without halts, voice breaking or other imperfections) were chosen for the study, with the rider that overall f_0 was roughly equated between languages. The chosen stimuli were manipulated to equate amplitude (at 65 dB) within and between language in PRAAT (Boersma & Weenink, 2013), but not duration to keep the stimuli sound natural. Among the four language systems, only Thai tones share similar durations across tones (due to vowel length being phonological); durations differ across tones in each of the other three languages.

2.3. Procedure

The task was run using the DMDX experimental platform (Forster & Forster, 2003). AX trials consisted of two syllables from the same language differing only in tone. The A-X inter-stimulus interval was 1000msec, and there were four types of tone pairs (Same: AA, BB; Different: AB, BA). To encourage phonemic rather than acoustic processing, in the same trials the two instances of the same tone were different exemplars of that tone (i.e., (A_1A_2, B_1B_2); and in the different trials (AB, BA) the tone exemplars in A and B positions differed between any particular set of trials. The total number of AX trials was 888, given by: [37 different tone pair trials] * [2 identity (same/different) * 2 order (AA vs BB or AB vs BA orders)] * [2C (/p, p^h/) × 3 V (/a, i, u/) = 6 syllable contexts]. To avoid fatigue effects, half the participants in each language group (12 out of 24) were tested on the 444 trials in which /p/ was the initial consonant and the other half on the 444 trials on which /p^h/ was the initial consonant. Participants were required to press the left/right shift key on a computer keyboard if they perceived the two sounds (AX) to be the same/different as quickly and accurately as possible.

2.4. Data treatment and analyses

CV contexts (/p, p^h/ x /a, i, u/) were included simply to enhance generalizability and, as they are of no specific interest, data were collapsed across CV contexts to provide greater power. To control for response biases, the dependent variable was d', defined as: *d' = Z(hit rate, where hit = 'different' response on an AB or BA trial) – Z (false alarm rate, where false alarm = 'different' response on an AA or BB trial)*.

3. Results

The data were submitted to linear mixed-effects modelling (LME) analyses using the lme4 (Bates et al., 2015) package's *lmer* function in R (R Core Team, 2018; R Studio Team, 2015). There were three key factors: listeners' language background (5 levels), tone system (4 levels) and contrast type (5 levels) and the random factor was participant. Results indicated that all three key factors played a role in listeners' overall performance. Post-hoc effects were obtained through pairwise comparisons using *emmeans.* Tone pair order across the various contrast types was characterized as 'high-to-low' versus 'low-to-high' as follows: For S-S trials, 'low-to-high' was defined as lower pitch height first and higher pitch height second, and vice versa for 'high-to-low'; for S-Dr/S-Df trials 'low-to-high' was defined as S as the first sound in the pair and D the second and vice versa for 'high-to-low' trials; for Df-Dr trials 'low-to-high' was defined as Df (decreasing pitch) as the first sound and Dr (increasing pitch) as the second and vice versa for 'high-to-low' trials; and for Dr-Dr trials 'low-to-high' was defined as the rising tone with the lesser dynamic contour as the first sound and that with the greater dynamic contour as the second and vice versa for 'high-to-low' trials. See Supplementary Materials for tables reporting the full model output.

3.1. Discrimination Performance: Cross-language tone perception

Fig. 1 shows the d' results for each of the four tone systems for each of the five participant groups. Linear mixed-effects modelling analyses revealed that tone language speakers performed significantly better than the non-tone language speakers (*estimate* = 0.518, standard error (*SE*) = 0.0946, z = 5.473, p <.0001). This effect was evident for contrasts in Cantonese, Thai and Chinese Mandarin tone systems (*estimates* > 0.351, *SEs* < 0.142, zs > 2.908, *ps* < 0.0037), but not for the Singaporean Mandarin tone system (*estimate* = 0.241, *SE* = 0.141, z = 1.713, p =.0868) in which scores across participants were high, a 'ceiling' effect.

Tone language listeners generally discriminated native tone contrasts significantly better than non-native contrasts (*estimate* = 0.374, *SE* = 0.0521, *z* = 7.178, p <.0001) (Fig. 2). This was evident for contrasts in the Cantonese, Thai and Chinese Mandarin (*estimates* > 0.2991, *SEs* < 0.1024, *zs* > 2.922, *ps* < 0.0036), but not the Singaporean Mandarin tone system (*estimate* = 0.0258, *SE* = 0.1255, *z* = 0.205, p =.8374) where participants across tone language background exhibited a ceiling effect.

Over and above modulations due to listeners' language background, discriminability of the four tone systems (Fig. 1A-1D) differed significantly (*estimates* > 0.257, *SEs* < 0.0648, *zs* > 4.457, *ps* < 0.0001). Cantonese tones (1A) were the most difficult to discriminate (*emmean* = 2.61, *SE* = 0.0477), with best discrimination by the native (dotted-line bar) Cantonese background listeners, (*ps* < 0.0066), and the worst by the non-tone language Australian English listeners (leftmost bar) (*ps* < 0.0180). This is followed by the Thai tones (1B) (*emmean* = 3.04, *SE* = 0.0519) with the Australian English listeners performing significantly worse than the native Thai (p =.0025) and the non-native Cantonese (p =.0134) listeners, although there was no difference between Thai and

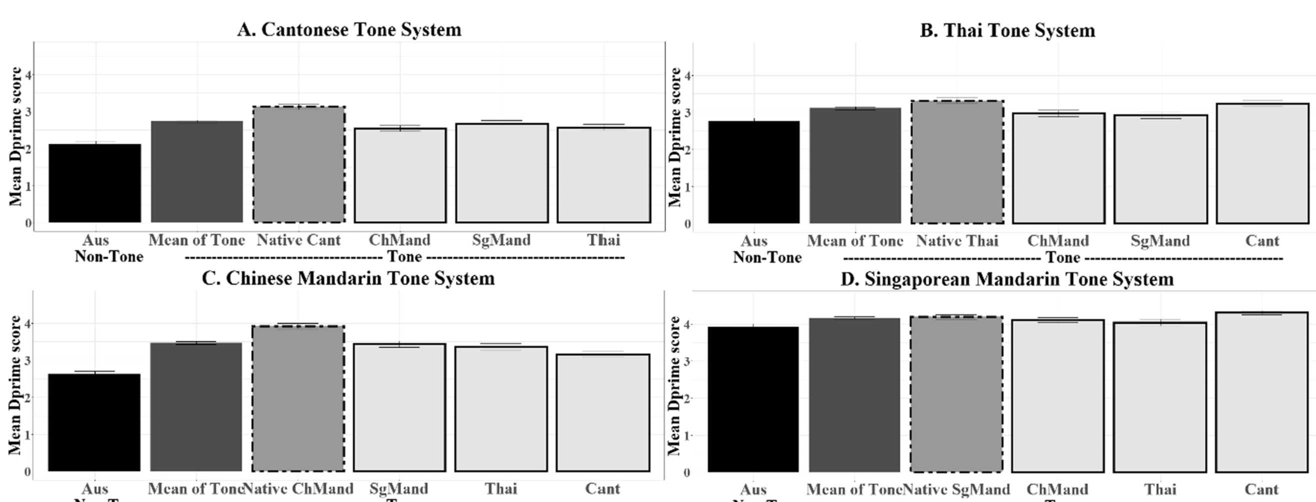

Fig. 1. Tone Language System and Language Background. For the four tone systems (a. Cantonese, b. Thai, c. Chinese Mandarin, d. Singaporean Mandarin), mean d' scores (y-axis) represent language background listeners' performances – across the x-axes, from left-to-right: the non-tone Australian English listeners, the mean performance across tone language speakers, then native speakers of that particular tone system (bar enclosed by a discontinuous line), then each of the tone language speakers listening to non-native tones. Error bars = ±1 SE.

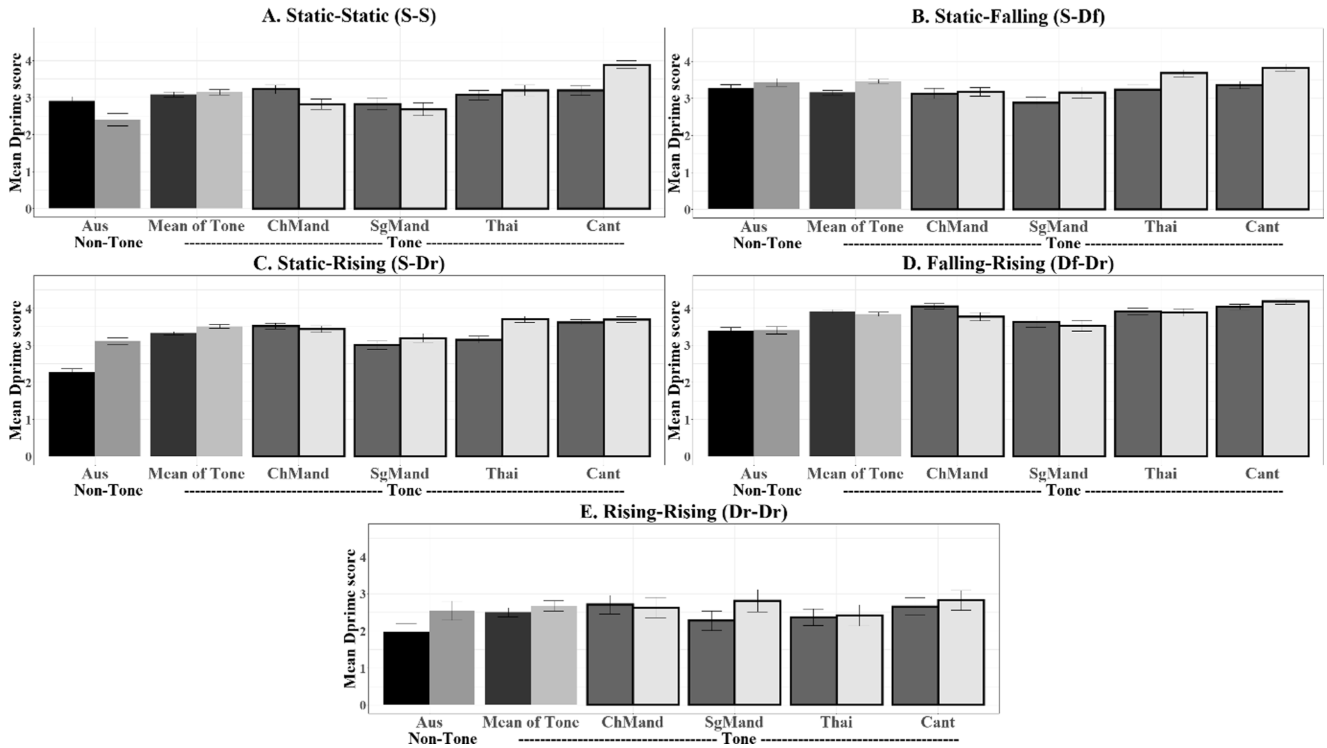

Fig. 2. Tone Contrast Types and Tone Contrast Orientation. For the five contrast types (a. Static-Static, b. Static-Falling, c. Static-Rising, d. Falling Rising, e. Rising-Rising), mean d' scores (y-axis) represent listeners' performances – across the x-axes, from left-to-right: the non-tone Australian English listeners, then mean performance across tone language speakers, then tone language speakers. Bar colours represent whether tone pairs were presented in the order from low-to-high (darker) or high-to-low (lighter) dynamic orientation. Error bar = ±1 SE.

Cantonese listeners (p =.9896). Discrimination of Chinese Mandarin tones (1C) was easier than either the Cantonese (p <.0001) or the Thai (<0.0001) tones (*emmean* = 3.30, *SE* = 0.0595), with the best performance by the native Chinese Mandarin listeners (*ps* < 0.0467) and the worst by the non-tone Australian English listeners (*ps* < 0.0211). Singaporean Mandarin tones (1D) were the easiest to discriminate (*emmean* = 4.11, *SE* = 0.0595) with no significant differences between listener groups (*ps* > 0.1805).

Across listener groups, discrimination performance differed between tone contrast types (Fig. 2) (*estimates* > 0.2761, *SEs* < 0.0836, *zs* > 4.739, *ps* < 0.0001). Df-Dr (2D) contrasts were the easiest to discriminate (*emmean* = 3.62, *SE* = 0.0583), then S-Df (2B) (*emmean* = 3.11, *SE* = 0.0561) and S-Dr (2C) (*emmean* = 3.07, *SE* = 0.0498) (which did not differ from each other (*estimate* = 0.0426, *SE* = 0.0531, z = 0.802, p =.9301), then S-S (2A) (*emmean* = 2.80, *SE* = 0.0611) with the Dr-Dr (2E) contrasts, being the most difficult (*emmean* = 2.37, *SE* = 0.0778). In terms of the specific native language influence, Cantonese listeners marginally outperformed Thai (p =.0685), Chinese (p =.0540), and significantly outperformed Singaporean (p =.0350) and Australian listeners (p =.0001) in S-S discrimination, and Australian English listeners underperformed other listeners (*ps* < 0.0227) in S-Dr and Df-Dr discrimination.

3.2. Discrimination Performance: Perceptual asymmetry among contrast types

Fig. 2 shows the d' scores for each of the four Tone Contrast Types and the two Tone Pair Orders for each of the five Language Background groups. Tests for orientational asymmetries for contrast type revealed that there were no overall cross-language background differences in the effect of order for the Df-Dr or S-S tone contrast types (*estimates* < 0.0509, *SEs* > 0.0746, *zs* < 0.681, *ps* > 0.4967). There were, however, significant overall cross-language background differences for: (i) Dr-Dr (more-to-less rising (lighter bars) better than less-to-more rising (darker bars), (ii) S-Df (Df → S pairs (lighter bars) better than S → Df pairs (darker bars); and (iii) S-Dr (Dr → S pairs (lighter bars) better than S → Dr pairs (darker bars) (*estimates* > 0.2521, *SEs* < 0.1142, *zs* > 2.209, *ps* < 0.0272). Within these three contrast types showing significant directional differences: (i) for Dr-Dr, there were no significant interactions with language background, i.e., the more-to-less rising (lighter bars) > less-to-more rising (darker bars) effect was consistent across language background groups, (ii) for S-Df, the Df → S pairs (lighter bars) > S → Df pairs (darker bars) effect was greater for Cantonese and Thai than the for other participants (*estimates* > 0.4382, *SEs* < 0.157, *zs* > 2.804, *ps* < 0.0051), and (iii) for S-Dr, the Dr → S pairs (lighter bars) > S → Dr pairs (darker bars) effect was greater for Australian English and Thai than other participants (*estimates* > 0.5520, *SEs* < 0.122, *zs* > 4.503, *ps* < 0.0001).

Taken together, in contrast types where differences occur, there is a general trend that discrimination of tone pairs is better when the first tone is more dynamic (higher pitch, greater pitch contour, or greater increase in pitch contour) than the second tone compared to when the first tone is less dynamic than the second tone.

3.3. Modelling

For MDS analyses, the d' scores were further fitted by the Individual Differences Scaling INDSCAL model (Carroll & Chang, 1970), a 3-way extension of the classical MDS using the *smacofIndDiff* function in the smacof R package (Borg, Groenen, & Mair, 2013; De Leeuw & Mair, 2009). A two-dimensional solution (pitch height, pitch direction) was opted for based on previous tone perception literature, visual inspection of the screen plots, and the interpretability of the dimensions. Both overall and language-specific cue weighting by listeners is reported using Kruskal's stress (Stress-1) to evaluate the model fit (Kruskal & Wish, 1978). Results with *randomstress()* baseline comparisons indicated a satisfactory model fit for four models corresponding to each language system (Mair, de Leeuw, & Groenen, 2015). The dimensions of each tone language were interpreted by matching each tone's f_0 and its derivatives (slope, acceleration etc., see Appendix III for detailed acoustic measurements) to the two dimensions of each tone system as closely as possible as follows: The acoustic measurements that were the most similar to the dimension coordinates were first selected (similarity measured by the Pearson's r correlation coefficient), and then confirmed by a set of regression analyses to ensure they were significant factors in predicting each dimension (height, direction). These height and direction weights were interpreted to index the relative importance a participant placed on the two cues in tone perception in a tone system and were compared within each INDSCAL model/tone system. Depending on the norming of the common space (of a tone system), they were considered as ranked data (Borg et al., 2013) and analysed by an ordinal regression model: Cumulative Link Mixed Model in R. Model assumptions were checked with the ordinal packages *nominal_test()* and *scale_test()* functions (Christensen, 2015). If any independent variable failed these tests (i.e., a significant p-value was returned), that variable was handled differently in the model using the nominal and scale options in the *clm()* function.

3.3.1. Modelling: Dimensionality

The INDSCAL model (Carroll & Chang, 1970) revealed pitch height and direction cues for the four tone systems (Fig. 3). The most salient height and direction cues for listeners were (i) for Thai tones (*stress-1* = 0.134): average pitch (r = 0.99, p =.001) and mean slope after the first inflection (r = 0.90, p =.038); (ii) for Cantonese tones (*stress-1* = 0.168): start pitch (r = -0.97, p =.001) and end slope (r = 0.96, p =.002); (iii) for Chinese Mandarin tones (*stress-1* = 0.090): end pitch (r = 0.97, p =.030) and mean magnitude of slope change rate over the last two time points (r = -0.99, p =.010); and (iv) for Singaporean Mandarin tones (*stress-1* = 0.077): min pitch (r = 0.99, p =.010) and end slope (r = -0.94, p =.060).

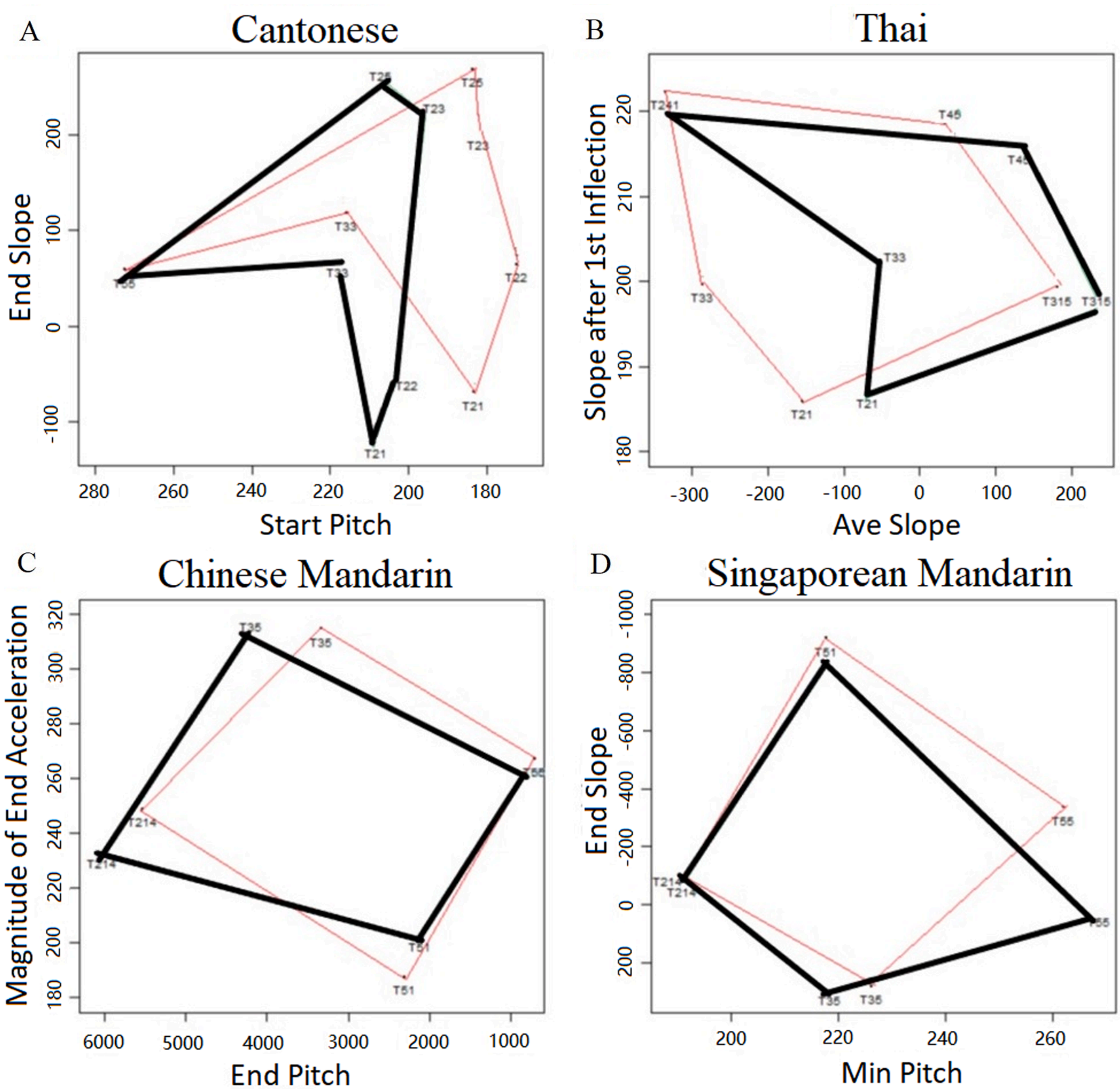

Fig. 3. Dimensionality modelling (INDSCAL schematics) of the four tone systems: (a. Cantonese, b. Thai, c. Chinese Mandarin, d. Singaporean Mandarin). The connected points represent acoustic space across listeners (black line) and interpreted dimensional solutions (red line) along the best fit height (x-axis) and direction (y-axis) cues most correlated to the two perceptual dimensions in each of the languages. The greater the overlap of the black and red spaces, the higher the likelihood that the interpreted perceptual dimensions are correct. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3.2. Modelling: Cue weighting

The main effects of language background and cue dimensions in each of the four tone systems were examined (Fig. 4 and Table 1, see Supplementary material for detailed weighting magnitudes across tone systems and language backgrounds). Overall, when tone background participants listen to tones in their native languages (dotted-line bar) Cantonese listeners favoured direction (End Slope) over height (Start Pitch) (D > H); Chinese Mandarin listeners favoured height (End Pitch) over direction (End Acc (abs) (H > D); Singaporean Mandarin listeners favoured direction (End Slope) over height (Min Pitch), and for Thai listeners, AvePitch and SlopeAfterInf were the most salient, but there was no bias for one or the other (D = H). These native language biases did not transfer directly to non-native tone languages, indeed there were many instances in which listeners adapted to the target language, e.g. when Cantonese speakers with their native bias of D > H listened to Chinese Mandarin, their bias changed to that of the native Chinese Mandarin speakers, H > D. There is, in fact, more adaptation to the L2 than there is L1 → L2 carryover bias. Most notable is that the non-tone language speakers did not adapt at all, they maintained an H > D bias, suggesting that experience with a particular tone language facilitates further facility with other tone languages.

4. Discussion

This study of the perception of four tone systems (Cantonese, Thai, Chinese Mandarin, Singaporean Mandarin) by listeners from five language backgrounds (Australian English, Cantonese, Thai, Chinese, Singaporean) showed that tone perception is constrained by (i) language background, (ii) tone system, (iii) tone contrast type, (iv) order of presentation, and (v) salience weighting of height and direction cues. Details of these findings are set out below.

4.1. Language background

Language background is a driving factor in tone perception. Consistent with previous literature (Burnham & Francis, 1997; Burnham, et al., 2015; Kaan, Wayland, Bao, & Barkley, 2007; Maggu, Zong, Law, & Wong, 2018), tone language adults show enhanced discrimination of tone compared with non-tone language adults. This facilitation, also evident in infancy research (Mattock & Burnham, 2006), was consistent across four of the five tone systems but was diminished in Singaporean Mandarin, which was found to be highly discriminable (see tone systems below).

While previous studies have indirectly shown support for the hypothesis that tone language listeners discriminate their native tones better than the non-native ones by comparing the perceptual outcomes of the same tones/tonal contrasts from different tone language speakers (e.g., Reid et al., 2015), our results show directly that tone language speakers' perception of native tones is better than their perception of non-native tones and do so quite convincingly – for three of the four tone language speaker groups here – Cantonese, Chinese Mandarin, and Thai (with the lack of effect for Singaporean Mandarin speakers being possibly due to all tone language speakers doing well on Singaporean tones). These observations likely point to the effect of specific linguistic experiences, such as perceptual assimilation be it on a phonetic (So & Best, 2014) or phonological (Bohn, Avesani, Best, & Vayra, 2011) level. These findings also dovetail with existing neurophysiological evidence for tone perception. In a positron emission tomography study comparing Thai and non-speech tones, Thai listeners showed significant activation in the left frontal operculum, suggesting phonological processing, whereas Chinese and English listeners exhibited sensitivity in the anterior insular region, indicating phonetic or non-linguistic processing. Hence, tone processing is constrained by language (tone versus non-tone; native versus non-native) experiences (Gandour et al., 2000).

4.2. Tone system

Discriminability of the four tone systems differed, with an ordinal ranking in discrimination ease, of Singaporean Mandarin (4 tones) > Chinese Mandarin (4 tones) > Thai (5 tones) > Cantonese (6 tones). At first glance, such ranking may be related to the complexity/density of tone space. However, that explanation cannot account for the difference between the two Mandarin tone systems which should share similar complexity. A recent study has proposed that Singaporean Mandarin tones may have a unique articulatory origin. Using ultrasound to track laryngeal movement, Yun and Moisik (2019) found that compared to Chinese Mandarin speakers, Singaporean Mandarin speakers' articulation of tones involves the additional use of laryngeal medialization and approximated thyroid lamina. It needs to be investigated whether these mechanisms also facilitate tone perception. The advantage for Singaporean Mandarin over Chinese Mandarin tones was evident in Australian English, Cantonese and Thai speakers, but the relative superiority was not so great for the Cantonese and Thai speakers. This may reflect similar static and dynamic tonal distributions/spaces in Chinese

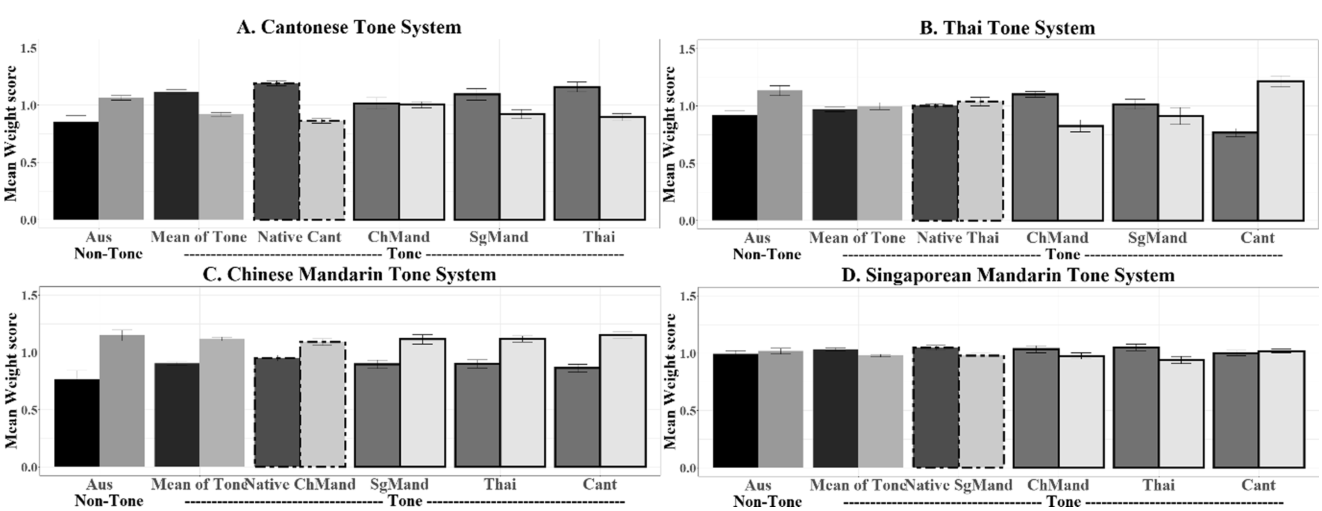

Fig. 4. Cue weighting scores for the four tone systems (a. Cantonese, b. Thai, c. Chinese Mandarin, d. Singaporean Mandarin). Mean weight scores (y-axis) represent listeners' performances (x-axis, left-to-right: Australian English listeners, then mean performance across tone language speakers, then native speakers of that particular tone system, then tone language speakers listening to non-native tones. Bar colours represent direction (darker) versus height (lighter) cues. Error bar = ±1 SE.

Table 1
Summary of the weighting of relative height (H) and direction (D) cues of the **4 tone systems** and the **5 language experience groups plus means for tone and non-native tone groups**, along with the specific significant **height and direction cues** with estimates (est), standard errors (SE) and z scores (* p <.015; ** p <.001; *** p <.0005) and gloss, i.e., whether D > H, H > D or D = H.

<table>
<tr><th rowspan="2">Tone System</th><th colspan="7">Language Background</th></tr>
<tr><th>Cantonese</th><th>Chinese Mandarin</th><th>Singaporean Mandarin</th><th>Thai</th><th>Australian (Non-Tone)</th><th>All Tone Languages</th><th>All Non-Nat Tone Langs</th></tr>
<tr><td>Cantonese</td><td>D>H
End Slope > Start Pitch
est = -2.59
SE = 0.300,
z = -8.652**</td><td>D=H
NSDs</td><td>D>H
End Slope > Start Pitch
est = -2.59
SE = 0.300,
z = -8.652**</td><td>D>H
End Slope > Start Pitch
est = -2.59
SE = 0.300
z = -8.652**</td><td>H>D
Start Pitch > End Slope
est = 1.7810 SE = 0.509
z = 3.497***</td><td>D>H</td><td>D>H</td></tr>
<tr><td rowspan="3">Ch-Mand</td><td>H>D</td><td>H>D</td><td>H>D</td><td>H>D</td><td>H>D</td><td rowspan="3">---</td><td rowspan="3">H>D</td></tr>
<tr><td></td><td>Ch-Mand vs rest
est = -1.787
SE = 0.493
z = -3.624***</td><td></td><td></td><td></td></tr>
<tr><td colspan="5">All 5: End Pitch > End Acc (abs), est > -1.786, SEs < 0.579 zs > -3.624**</td></tr>
<tr><td rowspan="2">Sg-Mand</td><td>D=H</td><td>D>H</td><td>D>H</td><td>D>H</td><td rowspan="2">D=H
NSDs</td><td rowspan="2">D>H</td><td rowspan="2">---</td></tr>
<tr><td colspan="4">End Slope > Min Pitch, ests > 1.214, SEs < 0.511, zs > -2.464*</td></tr>
<tr><td>Thai</td><td>H>D
AvePitch > SlopeAfterInf est = -2.324
SE = 0.540
z = -4.302**</td><td>D>H
SlopeAfterInf > AvePitch
est=3.065,
SE = 0.525
z = -5.834**</td><td>D=H
NSDs</td><td>D=H
NSDs</td><td>H>D
AvePitch > SlopeAfterInf est = -2.324 SE = 0.540
z = -4.302**</td><td>D=H</td><td>---</td></tr>
</table>

Mandarin and Singaporean Mandarin tone systems, and is also in line with findings showing the number of static tones in listeners' L1 can predict their discrimination ability at least for static tones (Chiao et al., 2011; Qin & Mok, 2011). Thus, there is evidence for a general constraint that the density/diversity of tone space in the native language affects tone discriminability irrespective of the listeners' language background. While this might be a key constraint affecting tone discriminability, it is not the only factor.

4.3. Tone contrast type

In order from easier to more difficult, the discrimination ease of the five contrast types was: Dynamic-Dynamic (Df-Dr) > Static-Static (S-S) > Static-Dynamic (S-Df = S-Dr) > Dynamic-Dynamic (Dr-Dr), which is in line with previous studies (Burnham et al., 1992; Burnham & Francis, 1997; Huang & Johnson, 2010; Liu & Kager, 2018). Turning to the effect of language background on contrast type, for Df-Dr, Australian English listeners performed more poorly than all tone language listeners, whose performance did not differ. For the most difficult Dr-Dr (rising versus. rising) contrast type, there was no influence of language background on performance, which could be interpreted as being due to the acoustic/ phonetic similarity of the two rising tones. However, it is equally plausible that for non-tone language speakers the difficulty is at the acoustic/phonetic level (So & Best, 2014), but for the tone language speakers, the difficulty may be more phonological (Reid et al., 2015). In support of such a conclusion, (i) the non-tone (Australian English) language listeners perform much better on the high-to-low order of presentation of Dr-Dr contrasts, whereas for the tone language background speakers this is not so apparent (greater acoustic effect for the non-tone language speakers), and (ii) Thai is the only tone language system which does not have a Dr-Dr contrast (see Supplementary material I) and Thai language background speakers perform more poorly on Dr-Dr contrasts than do the other three tone language background groups.

4.4. Tone order

The order in which tones are presented affected discriminability. Better discrimination of tone pairs was observed when the first tone is more dynamic (higher pitch, greater pitch contour, or greater increase in pitch contour) than the second tone compared to the other way around. Five explanations that have been proposed to account for such perceptual asymmetries can be considered. First, the under-specification hypothesis (Cornell, Lahiri, & Eulitz, 2013; Lahiri & Reetz, 2010) has been used to explain asymmetries in consonant (Gaskell, 2003; Hestvik & Durvasula, 2016), vowel (De Jonge & Boersma, 2015; Scharinger et al., 2012), and tone perception (Politzer-Ahles et al., 2016). This hypothesis argues that underspecified (and therefore more inclusive, flexible) features are better discriminated than specified features (Shafer, Schwartz, & Kurtzberg, 2004; Schluter et al., 2016; Schluter, Politzer-Ahles, Al Kaabi, & Almeida, 2017), resulting in the prediction that a presentation order with high-to-low dynamicity would be more underspecified. Second, in prototypicality theories, it is considered that non-native phonemes that are more deviant from prototypical members of a phonological category should be more salient and more easily discriminated from the prototype, than those closer to the prototypical members (Best, 1994; Best & Tyler, 2007; Ikeda et al., 2002; Kriengwatana & Escudero, 2017). This would imply that the more dynamic a tone the more it is prototypical. Third, in the Natural Referent Vowel framework

(Polka & Bohn, 2003, 2011), vowels that are more peripheral in F1-F2 space are discriminated better from more central vowels than vice versa, leading to the prediction that tones with higher dynamicity are considered more central in this model. The fourth explanation targets the psychophysics of tone salience (Burnham & Singh, 2018; Chandrasekaran, Krishnan, & Gandour, 2007, 2009; Krishnan, Gandour, & Bidelman, 2010) and is grounded in empirical findings demonstrating relationships between acoustic–phonetic properties and phonological development (Polka & Bohn, 2011; Tsuji, Mazuka, Cristia, & Fikkert, 2015). This would indicate an initial higher salience would be more distinguishable than if the initial tone were less salient. Finally, as pitch is a salient linguistic feature humans are exposed to since and before birth, we hypothesize that attentional factors may also play a role, such that orders leading to higher attention and arousal will be better discriminated, in our case the order of high-to-low dynamicity. Table 2 summarizes these explanations in the current context. These data that we report are the first systematic demonstration of tone asymmetry; contrasts presented from a high-to-low dynamic direction are better discriminated than those presented from a low-to-high dynamic direction. We further hypothesize this as a potential general constraint in tone discrimination and learning.

4.5. Cue dimensionality and weighting

The results regarding pitch height and direction cues reflect both flexibility and constraint, illustrating the tight bond between listeners' tone perception and the relative weight of height and direction cues in their native tone language (Francis et al., 2008). Of great importance are (i) differences in cue weighting between languages and (ii) how those differences may influence perceptual cue weighting in L2 tone perception.

First, there were systematic differences in how listeners from different language backgrounds weigh perceptual cues (Table 1). In the perception of their native tone systems, Chinese Mandarin listeners weighted height over direction cues whereas Singaporean Mandarin listeners weighted direction over height. Neither of these Mandarin varieties has two tones that differ only in f_0 height, leaving the unexplained finding of Chinese Mandarin listeners' greater reliance on height cues a subject for further exploration. Unlike Mandarin, in addition to contour tones, Cantonese and Thai each has three level tones contrasting in f_0 height with little f_0 modulation across time (Abramson, 1962; Morén & Zsiga, 2006), so the equal weighting of height and direction by native Thai speakers for Thai, and of direction over height by native Cantonese speakers for Cantonese are relatively explicable (although given the number of static: dynamic tones in each, 3:3 for Cantonese and 3:2 for Thai, the opposite could have been expected).

Second, regarding cues for L2 tone perception, previous studies indicate that Cantonese listeners show more reliance on direction than height cues (Gandour 1983; Li & Shuai, 2011), and that tone and non-tone language speakers attend to both height and direction cues when listening to Thai (Chandrasekaran et al., 2010). Our results depict a somewhat different picture (Table 1). *Chinese Mandarin listeners*, when listening to Singaporean Mandarin, shifted their native tone H > D reliance to align with that favoured by native Singaporean listeners, D > H; but while their reliance also changed for Thai (to D > H) and Cantonese (D = H), neither aligned with the native speakers' reliance. *Singaporean Mandarin listeners* exhibited perfect adaptability; given their D > H for native tones, they remained with D > H for Cantonese, shifted to H > D for Chinese Mandarin, and to H = D for Thai, weightings that might be considered optimal as they are the weightings that native listeners use for those languages (shaded cells in Table 1). *Cantonese listeners* were more similar to Chinese Mandarin listeners; they changed their native D > H weightings for all three of the other languages, but this only accorded with the native weighing in the case of Chinese Mandarin (H > D). Finally, Thai listeners showed some degree of adaptation: from their native D = H, they shifted to native language-appropriate weightings for Cantonese (D > H) and Chinese Mandarin (H > D), but not for Singaporean Mandarin (D > H rather than D = H).

The non-tone Australian English listeners showed strikingly different results. They were insensitive to the optimal cue weighting in different tone languages and consistently relied on H > D across the tone systems (except for Singaporean Mandarin – H = D). This is in line with previous studies on English listeners' reliance on height cues in their perception of Cantonese (Francis et al., 2008; Gandour, 1983; Gandour & Harshmen, 1978) and Thai (Burnham & Francis, 1997).

Overall, these results depict a slightly different picture from previous cue weighting studies which suggest that listeners' cross-language tone perception is determined by the cue weighting in their native language (e.g., Francis et al., 2008). Rather, it appears that tone language speakers are relatively flexible, shifting cues to a strategy that resembles the cue weighting of native speakers. Singaporean Mandarin and Thai listeners show the greatest adaptability, and their Cantonese and Chinese Mandarin counterparts do so with varying success. In contrast, the non-tone language listeners appeared insensitive to the differences between tone languages; they maintained an H > D strategy (H = D for Singaporean Mandarin) across tone languages.

Therefore, rather than a transfer of cue weighting priority from tone language speakers' L1 to L2, our results suggest that tone language speakers are sensitive to differences in tone systems, and that *it is the nature of the tone system rather than listeners' native tone language background that governs cue weighting in tone perception*. That said, having a tone language background is crucial for such adaptability. Australian English (non-tone) listeners were resistant and nonadaptive; they consistently weighted height over direction cues. This can be viewed in two ways, as (i) insensitivity to the nuances of tone languages resulting in non-optimal performance, or (ii) maintaining a preferred approach that, as a general strategy, works reasonably well across all tone languages, and specifically, selective attention to the general height cues that are present in most tone contrasts and selectively restricted attention to cues that are not universally useful across tone contrasts.

In sum, tone language listeners do not necessarily use the same cues in native and non-native tone perception. In fact, they tend to adapt flexibly to the acoustics of the target language rather than fixedly transferring their L1 priorities. Non-tone language listeners, on the other hand, do not show such adaptation, for them it appears that all tone languages are equal, as in equally foreign.

Table 2
A summary of perceptual asymmetry explanations against the current finding. Under these explanations, the presentation orders from a high to low dynamicity would be considered as being more specified, more typical and central representation, with higher salience and arousal than the other way around.

Results/ Hypotheses	Tone type A	Tone type B	Discrimination performance
Empirical Results (here)	High dynamicity	Low dynamicity	Hi → Lo > Lo → Hi (dynamicity)
Under-specification	Specified	Underspecified	Spec → Under > Under → Spec
Prototypes	Typical	Atypical	Typical → Atypical > Atypical → Typical
NRV	Central	Peripheral	Central → Periph > Periph → Central
Salience	High Salience	Low Salience	Hi → Lo > Lo → Hi (salience)
Attention	High Attention	Low Attention	Hi → Lo > Lo → Hi (attention)

5. Conclusion

Some previous research supports the notion that tone perception is flexible, independent of native phonology, whereas other research suggests it is constrained by native phonology or perceptual asymmetries. Given that such mixed results could be due to stimulus and/or task factors, we examined listeners' cross-language tonal sensitivity to five types of contrasts in four tone systems (Cantonese, Thai, Chinese Mandarin and Singaporean Mandarin) using the same design across five laboratories. Language background, tone features- and task-specific effects all emerged. There are effects of native language background that can constrain tone perception in an L2, but over and above this, the particular features of the tones in the target language exert a powerful influence on tone discrimination. Moreover, tone- but not non-tone language perceivers are sensitive to differences between tone languages; their tone language experience results in a certain flexibility – they adapt their relative reliance on height and direction cues to one that is most optimal for the target language. With respect to task-specific effects, we hypothesize a potential constraint, namely that pairwise tone discrimination is better when the first tone is more dynamic (higher pitch, greater pitch contour, or greater within-tone increase in pitch contour) than the second tone compared to the other way around. Further investigation of this effect may allow the determination of specific processes involved in tone perception. Thus, the findings of this study provide empirical evidence for why certain tones and/or tone languages are more difficult to perceive than others and grist for determining the implications of such differences. More generally, due to its high degree of breadth, depth and power, this study provides a a tone atlas, a cross-language-background tone perception of cross-tone-language-system corpus that can serve as a reference guide to inform empirical studies of tone sensitivity, both retrospectively and prospectively, upon which theories, future experiments, and judgments of lexical tone salience can be based.

Statement of significance

Testing 37 contrasts from four tone languages by listeners from five language backgrounds, we found a relationship between language-general, language-specific and contrast-specific experience and the performance of listeners' pitch perception and the cues they adopt, including a systematic directional asymmetry unseen in previous studies. Our findings provide empirical evidence and implications for variations in tone perception, providing a tone atlas from which theories, experiments and judgments of tone salience can be drawn.

Declaration of Competing Interest

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

Acknowledgement

Project conception (dB), and project management and data collection by the sixth author, BK, at the MARCS Institute for Brain, Behaviour and Development in Sydney Australia were supported by Australian Research Council Discovery Project grants (DP0988201, DP110105123) to the final author, dB. Data collection in Hong Kong was supported by Dr. Stanley Ho Medical Development Foundation. Data collection at the National University of Singapore was supported by an ODPRT grant for research excellence to LS. LL's writing was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 798658 hosted by Center for Multilingualism across the Lifespan at the University of Oslo, financed by Research Council of Norway through its Centers of Excellence funding scheme grant agreement No. 223265. MK's writing was supported by the Basque Government through the BERC 2018-2021 program, by the Spanish Ministry of Science and Innovation through the Ramon y Cajal Research Fellowship, PID2019-105528GA-I00, and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010-S. We would like to thank Kay Wong for data collection in Hong Kong, Ms. Juthatip Duangmal and Ms. Nawasri Chonmahatrakul at MARCS-CILS NokHook BabyLab, Thammasat University for data collection in Thailand, Charlene Fu and Dilu Wewalaarachchi for data collection in Singapore; and Antonia Götz for discussions about analyses in R.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2022.105106.

References

Abramson, A. S. (1962). *The vowels and tones of standard Thai: Acoustical measurements and experiments*. Indiana University.

Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language & Speech, 21*, 319–325.

Antoniou, M., & Wong, P. C. M. (2015). Poor phonetic perceivers are affected by cognitive load when resolving talker variability. *The Journal of the Acoustical Society of America, 138*(2), 571–574.

Antoniou, M., & Wong, P. C. M. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *The Journal of the Acoustical Society of America, 139*(1), 271–278.

Barry, J. G., & Blamey, P. J. (2004). The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese. *The Journal of the Acoustical Society of America, 116*(3), 1739–1748.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words, 167*(224), 233–277.

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech perception and linguistic experience*, 171–206.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language Experience in Second Language Speech Learning: In honor of James Emil Flege, 1334*, 1–47.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3. 51. *Online: http://www.Praat.Org*.

Bohn, O.-S., Avesani, C., Best, C. T., & Vayra, M. (2011). Perceiving through the lens of native phonetics: Italian and Danish listeners' perception of English consonant contrasts. *International Congress of Phonetic Sciences, 17*, 336–339.

Borg, I., Groenen, P., & Mair, P. (2013). *Applied Multidimensional Scaling*. Heidelberg: Springer.

Burnham, D. K., Attina, V., Xu, N., & Kasisopa, B. (2011). Towards an optimal tone space. In *Psycholinguistic Representation of Tone (PLRT) Conference: 22-23 August 2011, Hong Kong*.

Burnham, D., Brooker, R., & Reid, A. (2015). The effects of absolute pitch ability and musical training on lexical tone perception. *Psychology of Music, 43*(6), 881–897. https://doi.org/10.1177/0305735614546359

Burnham, D. & Francis, E. (1997) The role of linguistic experience in the perception of Thai tones. In A. S. Abramson (Ed) *SouthEast Asian Linguistic Studies in Honour of Vichin Panupong (Science of Language Vol. 8.)* Bangkok: Chulalongkorn University Press (29-47).

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Xu Rattanasone, N., Schwarz, I.-C., & Webster, D. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics, 77*, 571–591. https://doi.org/10.1017/S0142716414000496

Burnham, D., Kirkwood, K., Luksaneeyanawin, S., & Pansottee, S. (1992). Perception of Central Thai tones and segments by Thai and Australian adults. In *Pan-Asiatic linguistics: Proceedings of the third international symposium on language and linguistics* (pp. 546–560).

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. *Language experience in second language speech learning: In honor of James Emil Flege*, 259–280.

Burnham, D. K., & Singh, L. (2018). Coupling tonetics and perceptual attunement: The psychophysics of lexical tone contrast salience. *The Journal of the Acoustical Society of America, 144*(3), 1716–1716.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*(3), 283–319.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*, 148–156.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language, 108*(1), 1–9.

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

Chao, Y. R. (1930). A system of "tone letters". *Le Maître Phonétique, 30*, 24–27.

Chen, A., & Kager, R. (2016). Discrimination of lexical tones in the first year of life. *Infant and Child Development, 25*(5), 426–439.

Chen, A., Stevens, C. J., & Kager, R. (2017). Pitch perception in the first year of life, a comparison of lexical tones and musical pitch. *Frontiers in Psychology, 8*, 297.

Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of Mandarin tone sandhi. *Language Sciences, 48*, 62–69.

Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience, 31*(6), 751–760.

Chen, A., Peter, V., Wijnen, F., Schnack, H., & Burnham, D. (2018). Are lexical tones musical? Native language's influence on neural response to pitch in different domains. *Brain and Language, 180*, 31–41.

Chiao, W.-H., Kabak, B., & Braun, B. (2011).*When more is less: Non-native perception of level tone contrasts.* Retrieved February 2, 2012, from http://ling.uni-konstanz.de/pages/home/braun/articles/Chiao.Kabak.Braun-1.pdf.

Choi, W., Tong, X., & Singh, L. (2017). From lexical tone to lexical stress: A cross-language mediation model for Cantonese children learning English as a second language. *Frontiers in psychology, 8*, 492.

Christensen, R. H. B. (2015). Ordinal—regression models for ordinal data. *R package version, 28*, 2015.

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America, 131*(6), 4756–4769.

Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance, 39*(3), 757.

De Jonge, M. J., & Boersma, P. (2015). French high-mid vowels are underspecified for height. In *Proceedings of the 18th International Congress of Phonetic Sciences* (Glasgow: The University of Glasgow).

De Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software., 31*, 1–30.

Feng, G., Gan, Z., Wang, S., Wong, P. C. M., & Chandrasekaran, B. (2018). Task-general and acoustic-invariant neural representation of speech categories in the human brain. *Cerebral cortex, 28*(9), 3241–3254.

Feng, S., Wu, A., & Nissenbaum, J. (2020). Perceived pitch and formant frequencies in the perception of lexical tone in Cantonese. *Poster presented at the 50th Meeting of the North East Linguistic Society, MIT*.

Fikkert, P., Liu, L., & Ota, M. (2020). The Acquisition of Word Prosody. In *The Oxford Handbook of Language Prosody*.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116–124.

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.

Fromkin, V. (Ed.). (1978). *Tone: A linguistic survey*. New York: Academic Press.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11*, 149–175.

Gandour, J., Dzemidzic, 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.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., … Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage, 23*(1), 344–357.

Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., & Li, X. (2003). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Human Brain Mapping, 18*(3), 149–157.

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and speech, 21*(1), 1–33.

Gandour, J. T., & Krishnan, A. (2016). Processing tone languages. In *Neurobiology of language* (pp. 1095–1107). Academic Press.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Lancker, D. V., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of cognitive neuroscience, 12*(1), 207–222.

Gandour, J., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *NeuroReport, 9*(9), 2115–2119.

Gaskell, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics, 31*(3–4), 447–463.

Götz, A., Yeung, H. H., Krasotkina, A., Schwarzer, G., & Höhle, B. (2018). Perceptual reorganization of lexical tones: Effects of age and experimental procedure. *Frontiers in psychology, 9*, 477.

He, C., Hotson, L., & Trainor, L. J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of cognitive neuroscience, 19*(5), 878–892.

Hestvik, A., & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain and Language, 152*, 28–43.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge University Press.

Huang, T., & Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica, 67*(4), 243–267.

Ikeda, K., Hayashi, A., Hashimoto, S., Otomo, K., & Kanno, A. (2002). Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference. *Neuroscience Letters, 321*, 133–136.

Ingvalson, E. M., Barr, A. M., & Wong, P. C. M. (2013). Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning. *J. Speech Lang. Hear Res., 56*(3), 1045–1050.

Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain research, 1148*, 113–122.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35*(1), 104–117.

Kriengwatana, B. P., & Escudero, P. (2017). Directional asymmetries in vowel perception of adult nonnative listeners do not change over time with language experience. *Journal of Speech, Language, and Hearing Research, 60*, 1088–1093.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brain stem. *Journal of Neurolinguistics, 23*, 81–95.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling. Quantitative Applications in the Social Sciences*. Beverly Hills, CA: Sage.

Kwok, V. P., Dan, G., Yakpo, K., Matthews, S., & Tan, L. H. (2016). Neural systems for auditory perception of lexical tones. *Journal of Neurolinguistics, 37*, 34–40.

Kwok, V. P., Dan, G., Yakpo, K., Matthews, S., Fox, P. T., Li, P., & Tan, L. H. (2017). A meta-analytic study of the neural systems for auditory processing of lexical tones. *Frontiers in human neuroscience, 11*, 375.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics, 38*(1), 44–59.

Li, B., & Shuai, L. (2011, August). Effects of native language on perception of level and falling tones. In *ICPhS* (pp. 1202-1205).

Li, X., Gandour, J. T., Talavage, T., Wong, D., Hoffa, A., Lowe, M., & Dzemidzic, M. (2010). Hemispheric asymmetries in phonological processing of tones vs. segmental units. *NeuroReport, 21*(10), 690.

Liang, B., & Du, Y. (2018). The functional neuroanatomy of lexical tone perception: An activation likelihood estimation meta-analysis. *Frontiers in neuroscience, 12*, 495.

Lin, M. C. (1987). The perceptual cues of tones in standard Chinese. *Proceedings of the Eleventh International Congress of Phonetic Sciences, Vol. 1* (pp. 162–165). Tallinn, Estonia, U.S.S.R.: Academy of Sciences of the Estonian SSR.

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition, 133*(2), 385–394.

Liu, L., & Kager, R. (2017). Perception of tones by bilingual infants learning non-tone languages. *Bilingualism, 20*(3), 561.

Liu, L., & Kager, R. (2018). Monolingual and bilingual infants' ability to use non-native tone for word learning deteriorates by the second year after birth. *Frontiers in Psychology, 9*, 117.

Liu, L., Ong, J. H., Tuninetti, A., & Escudero, P. (2018). One way or another: Evidence for perceptual asymmetry in pre-attentive learning of non-native contrasts. *Frontiers in psychology, 9*, 162.

Liu, L., Chen, A., & Kager, R. (2020). Simultaneous bilinguals who do not speak a tone language show enhancement in pitch sensitivity but not in executive function. *Linguistic Approaches to Bilingualism, 3*.

Maggu, A. R., Zong, W., Law, V., & Wong, P. C. M. (2018). Learning two tone languages enhances the brainstem encoding of lexical tones. Proceedings of the Interspeech, September, 2018, Hyderabad, India.

Mair, P., de Leeuw, J., & Groenen, P. J. (2015). *Multidimensional scaling in R: smacof*. URL: https://cran.r-project.org/web/packages/smacof/vignettes/smacof.pdf.

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.

Masapollo, M., Zhao, T. C., Franklin, L., & Morgan, J. L. (2019). Asymmetric discrimination of nonspeech tonal analogues of vowels. *Journal of Experimental Psychology: Human Perception and Performance, 45*(2), 285.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241–265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*(3), 1367–1381.

Morén, B., & Zsiga, E. (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory, 24*(1), 113–178.

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics, 38*(4), 616–624.

Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance, 42*(10), 1547.

Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*(1), 221–231.

Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics, 39*(4), 467–478.

Polka, L., Ruan, Y. F., & Masapollo, M. (2018). Understanding vowel perception biases—It's time to take a meta-analytic approach. *Manuscript submitted for publication.*

Qin, Z., & Mok, P.-P.-K. (2011). Discrimination of Cantonese tones by Mandarin, English and French speakers. In *The psycholinguistic representation of tone, 2011* (pp. 50–53). Hong Kong: Causal Productions.

Quam, C., & Swingley, D. (2010). Phonological knowledge guides 2-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of memory and language, 62*(2), 135–150.

R Studio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

R Core Team (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramachers, S., Brouwer, S., & Fikkert, P. (2018). No perceptual reorganization for Limburgian tones? A cross-linguistic investigation with 6-to 12-month-old infants. *Journal of child language, 45*(2), 290.

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

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2012). Asymmetries in the processing of vowel height. *Journal of Speech, Language, and Hearing Research, 55*, 903–918.

Schluter, K., Politzer-Ahles, S., & Almeida, D. (2016). No place for/h: An ERP investigation of English fricative place features. *Language, Cognition and Neuroscience, 31*, 728–740.
Shafer, V. L., Schwartz, R. G., & Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cognitive Brain Research, 18*, 242–254.
Singh, L., Fu, C. S., Seet, X. H., Tong, A. P., Wang, J. L., & Best, C. T. (2018). Developmental change in tone perception in Mandarin monolingual, English monolingual, and Mandarin-English bilingual infants: Divergences between monolingual and bilingual learners. *Journal of Experimental Child Psychology, 173*, 59–77.
So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition, 36*, 195–221.
Tong, X., Lee, S. M. K., Lee, M. M. L., & Burnham, D. (2015). A tale of two features: Perception of Cantonese lexical tone and English lexical stress in Cantonese-English bilinguals. *PLoS ONE, 10*(11), e0142896.
Tsao, F. M. (2017). Perceptual improvement of lexical tones in infants: Effects of tone language experience. *Frontiers in psychology, 8*, 558.
Tsuji, S., Mazuka, R., Cristia, A., & Fikkert, P. (2015). Even at 4 months, a labial is a good enough coronal, but not vice versa. *Cognition, 134*, 252–256.
Vance, T. J. (1977). Tonal distinctions in Cantonese. *Phonetica, 34*, 93–107.
Wayland, R., & Chen, S. (2018, November). Asymmetries in lexical tone perception. In *Proceedings of Meetings on Acoustics 176ASA* (Vol. 35, No. 1, p. 060006). Acoustical Society of America.
Wayland, R., Chen, S., Zhou, F., & Hong, Y. (2019). Directional asymmetry in lexical tone perception. *Proceedings of Meetings on Acoustics, 39*, 060005.
Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics, 39*(4), 703–728.
Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter-and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research, 46*(2), 413–421.
Wong, P. C. M., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The Role of the Insular Cortex in Pitch Pattern Perception: The Effect of Linguistic Contexts. *Journal of Neuroscience, 24*, 9153–9160.
Wong, P. C. M., & Perrachione, T. K. (2007). Learning Pitch Patterns in Lexical Identification by Native English-Speaking Adults. *Applied Psycholinguistics, 28*, 565–585.
Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience, 170*(1), 223–231.
Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of phonetics, 25*(1), 61–83.
Yang, M., & Sundara, M. (2019). Cue-shifting between acoustic cues: Evidence for directional asymmetry. *Journal of Phonetics, 75*, 27–42.
Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68*(2), 123–139.
Yip, M. J. W. (2002). *Tone* (Chap. 1, pp. 1–14). New York: Cambridge University Press.
Yun, D. & Moisik, S. (2019). A laryngeal ultrasound study of Singaporean Mandarin tones. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (Eds.) Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019.
Zhao, T. C., Masapollo, M., Polka, L., Ménard, L., & Kuhl, P. K. (2019). Effects of formant proximity and stimulus prototypicality on the neural discrimination of vowels: Evidence from the auditory frequency-following response. *Brain and language, 194*, 77–83.
Zsiga, E., & Nitisaroj, R. (2007). Tone features, tone perception, and peak alignment in Thai. *Language and Speech, 50*(3), 343–383.