AUGUST 11 2023

Does musicianship influence the perceptual integrality of tones and segmental information?

William Choi ; Veronica Ka Wai Lai

J Acoust Soc Am 154, 852–862 (2023)
https://doi.org/10.1121/10.0020579

JASA ARTICLE

Does musicianship influence the perceptual integrality of tones and segmental information?

William Choi[1,a] and Veronica Ka Wai Lai[2]
[1]Academic Unit of Human Communication, Development, and Information Sciences, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region
[2]Department of Pediatrics, Faculty of Medicine, Manitoba University, Winnipeg, Manitoba, Canada

ABSTRACT:
This study investigated the effect of musicianship on the perceptual integrality of tones and segmental information in non-native speech perception. We tested 112 Cantonese musicians, Cantonese non-musicians, English musicians, and English non-musicians with a modified Thai tone AX discrimination task. In the tone discrimination task, the control block only contained tonal variations, whereas the orthogonal block contained both tonal and task-irrelevant segmental variations. Relative to their own performance in the control block, the Cantonese listeners showed decreased sensitivity index (d') and increased response time in the orthogonal block, reflecting integral perception of tones and segmental information. By contrast, the English listeners performed similarly across the two blocks, indicating independent perception. Bayesian analysis revealed that the Cantonese musicians and the Cantonese non-musicians perceived Thai tones and segmental information equally integrally. Moreover, the English musicians and the English non-musicians showed similar degrees of independent perception. Based on the above results, musicianship does not seem to influence tone-segmental perceptual integrality. While musicianship apparently enhances tone sensitivity, not all musical advantages are transferrable to the language domain.
 https://doi.org/10.1121/10.0020579
(Received 23 February 2023; revised 26 June 2023; accepted 21 July 2023; published online 11 August 2023)
[Editor: Melissa Michaud Baese-Berk] Pages: 852–862

I. INTRODUCTION

There are two fundamental questions in the study of speech perception: first, how humans perceive segmental (e.g., consonant and vowel) and suprasegmental (e.g., tone, intonation, and stress) speech information (Freeman, 2023; McClelland and Elman, 1986; Sinagra and Wiener, 2022), and second, whether and how humans combine different speech dimensions to form a unified mental percept (*perceptual integration*; Choi *et al.*, 2017; Garner, 1974; Yu *et al.*, 2022). In the music-to-language transfer literature, most studies pertain to the first fundamental question. For example, they have identified the positive effect of musicianship on tone perception (Choi, 2020; Zheng and Samuel, 2018), stress perception (Choi, 2022a; Kolinsky *et al.*, 2009), and segmental perception (Cooper *et al.*, 2017; Sadakata and Sekiyama, 2011). Based on empirical works, theoreticians have promisingly captured how musicianship enhances perceptual sensitivities to segmental and suprasegmental information (Patel, 2011, 2014; Tierney and Kraus, 2014). Comparatively, no study to our knowledge has attempted to address the second fundamental question (i.e., perceptual integration) in the context of music-to-language transfer. Addressing this research gap will inform cognitive psychologists of the need (or the lack thereof) to devise a theoretical hypothesis on how musicianship influences tone-segmental integrality. Taking the first step, the present study investigated whether musicianship influences tone-segmental perceptual integrality in Cantonese and English listeners.

A. Musicianship enhances perceptual sensitivities

The recent decade has seen emerging theoretical accounts of how musicianship enhances speech sensitivity (Patel, 2011; 2014; Tierney and Kraus, 2014). According to the OPERA hypothesis, music training that satisfies five conditions (overlap, precision, emotion, repetition, and attention) would enhance the subcortical encoding of the acoustic features of speech, thereby sharpening speech sensitivity [for descriptions of the five conditions, see Patel (2011, 2014)]. Focusing on the timing aspect of music and speech, the precise auditory timing hypothesis posits that music entrainment activities facilitate the detection of fine-grained timing details in speech, in turn improving consonantal and vocalic perception (Tierney and Kraus, 2014).

Indeed, empirical studies have identified robust links between musicianship and speech sensitivity. Relative to their non-musician counterparts, English and Italian musicians discriminated Mandarin tones more accurately (Alexander *et al.*, 2005; Delogu *et al.*, 2010). In a later

[a]Also at: Speech and Music Perception Laboratory, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region. Electronic mail: willchoi@hku.hk

study, English musicians were better able to detect tonal differences in Mandarin phrases than English non-musicians (Zheng and Samuel, 2018). This has reflected that the musical advantage also applies to the phrase level. A more recent study tested English musicians and non-musicians on Cantonese tone discrimination and sequence recall (Choi, 2020). In both tasks, the musicians outperformed the non-musicians on some but not all Cantonese tones, suggesting that their musical advantage was selective to some tones rather than general. In addition to tones, English and French musicians also discriminate and recall English stress sequences more accurately than their non-musician counterparts (Choi, 2022a; Kolinsky *et al.*, 2009). Segmentally, English musicians outperformed English non-musicians on discriminating phonemic length vocalic contrasts in Thai language (Cooper *et al.*, 2017). Subtle differences aside, the above studies on tone, stress, and vocalic perception collectively support the theoretical notion that musicianship enhances speech sensitivities.

B. Effect of first language experience on tone-segmental perceptual integrality

From a theoretical perspective, speech perception concerns not only perceptual sensitivities but also *how* segmental and suprasegmental information are processed. In the speech perception research field, a prominent question is how language experience drives listeners to integrally (or independently) perceive segmental and suprasegmental information (Algom and Fitousi, 2016). Previous studies have utilized the Garner paradigm (Lee and Nusbaum, 1993; Lin and Francis, 2014; Tong *et al.*, 2008; Yu and Zhang, 2018) and the mismatch negativity (MMN) additivity paradigm (Choi *et al.*, 2017; Gao *et al.*, 2012; Yu *et al.*, 2022) to address this question. As the present study is behavioral in nature, we review the behavioral evidence in greater detail.

The behavioral-based Garner paradigm typically comprises a speeded classification task with a control block and an orthogonal block (Garner, 1974). In the control block, only the target dimension (e.g., tones) varies. By contrast, the orthogonal block contains both target (e.g., tones) and non-target (e.g., vowels) variations. Known as Garner interference, the increase in response time in the orthogonal block relative to the control block indicates integral perception of the target and non-target dimensions [for a review, see Algom and Fitousi (2016)]. Based on the same principle, Choi and Tsui (2022) developed a more sensitive tool to probe Garner interference—the modified AX discrimination task. On each trial, listeners discriminate a target dimension in the presence (orthogonal block) or absence of non-target variations (control block). Unlike in the speeded classification task, Garner interference manifested not only as increased response time but also as decreased sensitivity index and accuracy (Choi and Tsui, 2022).

For tone language listeners, behavioral and neural evidence suggests that they perceive tone and segmental information integrally in their first language (e.g., Choi *et al.*, 2017; Repp and Lin, 1990; Yu and Zhang, 2018). An early study adopted a speeded classification paradigm in which Mandarin and English listeners classified consonants (/b/ or /d/), vowels (/a/ or /u/), and tones (high or falling) in the control and orthogonal conditions (Repp and Lin, 1990). In the control condition of the tone classification task, the two tones were embedded in the same syllable (e.g., /ba-high/ /ba-falling/ /ba-falling/ /ba-high/). In the orthogonal condition, the two tones were embedded in different syllables (e.g., /ba-high/ /bu-high/ /ba-falling/ /bu-falling/). The listeners were asked to ignore task-irrelevant changes (e.g., vowel) and attend only to the task-relevant information (e.g., tones). Relative to their own performance in the respective control blocks, the Mandarin listeners were slower in classifying tones, consonants, and vowels in the orthogonal blocks. In a more recent study, Mandarin and English listeners completed a speeded vowel classification task with the control blocks (e.g., /da-rising/ /du-rising/ /du-rising/ /da-rising/) and the orthogonal block (e.g., /da-rising/ /du-falling-rising/ /du-falling-rising/ /da-rising/) (Yu and Zhang, 2018). As in the early study (Repp and Lin, 1990), the Mandarin listeners were slower in the orthogonal block than in the control block. Furthermore, the Mandarin listeners' behavioral Garner interference effect correlated with their subcortical frequency-following responses for Mandarin tone contours, suggesting a native language neural commitment at the subcortical level [Yu and Zhang, 2018; see also Zhang *et al.* (2005)]. In addition to Mandarin listeners, MMN additivity studies have shown that Cantonese listeners perceive tones and vowels integrally (Choi *et al.*, 2017; Yu *et al.*, 2022). Collectively, behavioral and neurophysiological evidence has indicated that Cantonese and Mandarin listeners perceive tones and segmental information integrally (e.g., Choi *et al.*, 2017; Repp and Lin, 1990; Yu and Zhang, 2018).

Is tone-segmental integrality transferrable to non-native speech? The above studies only tested the tone language listeners on their first languages (e.g., Choi *et al.*, 2017; Repp and Lin, 1990; Yu and Zhang, 2018). Thus, it was unclear whether their perceptual integrality was transferrable across languages. In a modified AX discrimination task, Cantonese listeners discriminated Thai tones with and without segmental variations, i.e., orthogonal and control blocks, respectively (Choi and Tsui, 2022). In the orthogonal block, the Cantonese listeners showed increased response time, decreased sensitivity index, and decreased accuracy. These suggested that their tone-segmental perceptual integrality further applies to non-native tones and segments (Choi and Tsui, 2022).

For English listeners, behavioral studies have suggested that they perceive tones and segmental information independently (Choi and Tsui, 2022; Lin and Francis, 2014; Zou *et al.*, 2017; cf. Lee and Nusbaum, 1993; Repp and Lin, 1990). In a previous study, English listeners classified Mandarin tones and consonants with similar response time across the control and orthogonal blocks (Lin and Francis, 2014). In a modified Thai tone AX discrimination task,

segmental variation did not affect English listeners' response time, accuracy, and sensitivity index (Choi and Tsui, 2022). That said, Repp and Lin (1990) reported a Garner interference effect of vowel variation on tone classification in English listeners. Nevertheless, the English listeners experienced a weaker Garner interference effect than the Mandarin listeners. Taken together, the above studies have reflected that English listeners perceive tones and segmental information independently (e.g., Choi and Tsui, 2022; Lin and Francis, 2014) or at least less integrally than tone language listeners (e.g., Repp and Lin, 1990).

Based on the above findings, Choi and Tsui (2022) proposed the dimension transfer hypothesis of tone-segmental integrality. The hypothesis posits that tone-segmental integrality (or non-integrality) is transferrable to non-native speech. Although the Thai tones and segmental information are absent in Cantonese, Cantonese listeners still perceived them integrally in the previous study (Choi and Tsui, 2022). As such, the hypothesis postulates that the transfer is cognitively driven rather than phonetically driven. Specifically, the transfer does not involve the transfer of a particular L1 phoneme, tone, or acoustic cue, but rather the cognitive experience of integrating segmental and tonal information. In a similar vein, English listeners process English segmental and suprasegmental (tonal) information independently as separate information (e.g., Choi and Tsui, 2022; Lin and Francis, 2014) or at least less integrally than tone language listeners (e.g., Repp and Lin, 1990). Thus, their cognitive experience in not (or at least weakly) integrating segmental and tonal information transfers to non-native speech, e.g., Mandarin and Thai. In brief, the dimension transfer hypothesis posits that language experience shapes tone-segmental integrality in non-native speech. As detailed below, it is also possible that musicianship and language experience can jointly influence tone-segmental integrality in non-native speech (Hansen *et al.*, 2022; Neuhaus and Knösche, 2008; Prince, 2011). If that is indeed true, it would highlight the theoretical need to incorporate the music factor into the dimension transfer hypothesis.

C. Is the musicianship effect on perceptual integration transferrable to the language domain?

There are mixed findings on the effect of musicianship on perceptual integrality. On the one hand, ample evidence shows that musicianship drives listeners to process multiple music features integrally (Hansen *et al.*, 2022; Neuhaus and Knösche, 2008; Prince, 2011). In a behavioral study, musicians and trained non-musicians completed a goodness rating task (Prince, 2011). The participants heard melodic sequences with various degrees of conformity to the pitch and timing aspects of typical Western music (Berkowitz *et al.*, 1997; Ottman, 1986). On each trial, an explicit instruction required the participants to evaluate how well formed the melodic sequence was by attending only to the pitch (i.e., ignore time), time (i.e., ignore pitch), or both. Prince (2011) found that the musicians and non-musicians were unable to ignore time when rating pitch, and vice versa. Moreover, the musicians experienced a stronger interference effect than the non-musicians. These findings suggested that musicianship induces perceptual integrality, at least for pitch and time (Prince, 2011). This was consistent with a neurophysiological study that found that musicians perceived pitch and time more integrally than non-musicians (Neuhaus and Knösche, 2008). In addition to pitch and time, a recent MMN additivity study examined whether musicians and non-musicians perceived pitch, intensity, and perceived location of musical notes integrally (Hansen *et al.*, 2022). Relative to the non-musicians, the musicians showed more integrative perception of the three musical features. Taken together, the behavioral and neural findings suggest that musicianship induces perceptual integrality of music features (Hansen *et al.*, 2022; Neuhaus and Knösche, 2008; Prince, 2011). This allows musicians to efficiently combine multiple features to form complex representations (Hansen *et al.*, 2022). If the positive effect of musicianship on perceptual integrality applies to the language domain, musicianship should enhance tone-segmental perceptual integrality in Cantonese listeners. As such, Cantonese musicians should exhibit a higher degree of tone-segmental perceptual integrality than Cantonese non-musicians. Additionally, musicianship should induce tone-segmental perceptual integrality in English listeners. While English non-musicians perceive tonal and segmental information independently, English musicians should show integral perception.

On the other hand, there is as much evidence showing that musicianship drives listeners to process multiple music features independently (Ahissar and Hochstein, 2004; Ahissar *et al.*, 2009). Music training often involves analytical experience in breaking down various music elements, such as individual notes, instruments, rhythms, and harmonies. Thus, it is possible that musicianship inhibits perceptual integrality. A previous study showed that timbre variations interfered with pitch perception (Allen and Oxenham, 2014). Relative to non-musicians, musicians were less affected by timbre variations when discriminating pitch (Pitt, 1994), and vice versa (Melara and Marks, 1990). These findings have suggested that musicians show less integral processing than non-musicians. If the negative effect of musicianship on perceptual integrality applies to the language domain, musicianship should decrease tone-segmental perceptual integrality in Cantonese listeners. Thus, Cantonese musicians should perceive tonal and segmental information less integrally than Cantonese non-musicians. Even without music learning experience, English listeners are already perceiving tonal and segmental information independently (Choi and Tsui, 2022; Lin and Francis, 2014). Therefore, English musicians and non-musicians should not differ in terms of perceptual integrality.

Given the possible positive (Hansen *et al.*, 2022; Neuhaus and Knösche, 2008; Prince, 2011) or negative effect of musicianship (Ahissar *et al.*, 2009; Melara and Marks, 1990; Pitt, 1994) on perceptual integrality within the

https://doi.org/10.1121/10.0020579

music domain, this study investigates whether such an effect is transferrable to the language domain. In the music-to-language transfer literature, empirical and theoretical works have largely focused on perceptual sensitivities (Cooper *et al.*, 2017; Kolinsky *et al.*, 2009; Patel, 2014; Sadakata and Sekiyama, 2011; Tierney and Kraus, 2014). To our knowledge, there has been no attempt to address how musicianship shapes perceptual integrality in the language domain. The lack of relevant empirical data has inherently limited the theoretical work that could be done. Therefore, it is theoretically important to address this research gap. A positive or negative result will provide grist for a theoretical model pertaining to the joint effect of musicianship and first language experience on non-native speech perceptual integrality. A null result will inform theoreticians of the limitation of music-to-language transfer and the lack of need to devise the aforementioned model.

In short, this study aimed to investigate whether musicianship and first language experience jointly influence tone-segmental perceptual integrality in non-native speech. Regarding first language experience, it was established that Cantonese listeners perceive non-native tonal and segmental information integrally, whereas English listeners perceive them separately (Choi and Tsui, 2022). Concerning musicianship, there are three possible results. The first possibility is that musicianship increases perceptual integrality in Cantonese listeners and induces it in English listeners. The second possibility is that musicianship decreases perceptual integrality in Cantonese listeners and exerts no effect on English listeners. The third possibility is that musicianship does not affect perceptual integrality in Cantonese listeners and English listeners.

II. METHODS

A. Participants

The final sample contained 112 participants, including 30 Cantonese musicians, 25 Cantonese non-musicians, 27 English musicians, and 30 English non-musicians. All Cantonese listeners (i) had learnt and spoke Cantonese as their first language, (ii) had not learnt Thai, and (iii) reported having typical hearing. All English listeners (i) had learnt and spoke English as their first language, (ii) had not learnt any tone language, and (iii) reported having typical hearing. Based on pre-established criteria, musicians are individuals with (a) over 7 years of continuous formal music training and (b) self-reported ability to play one or more musical instruments (Choi, 2020, 2022b; Cooper and Wang, 2012). Non-musicians are individuals with (a) no more than 2 years of music training, (b) no formal music training in the past 5 years, and (c) self-reported inability to play any musical instruments. In any case, typical classroom-based school music lessons were not counted as music training (Miya, 2005).

All participants completed a music background questionnaire (Choi, 2021). On average, the Cantonese musicians received 10.8 years [standard deviation (SD) = 2.9 years] of music training starting from 7.8 years old (SD = 3.5 years), the Cantonese non-musicians received 0.3 years (SD = 0.6 years) of music training starting from 12.3 years old (SD = 6.9 years), the English musicians received 10.4 years (SD = 3.2 years) of music training starting from 9 years old (SD = 3.5 years), and the English non-musicians received 0.4 years (SD = 0.7 years) of music training starting from 11.6 years old (SD = 2.9 years). At time of testing, their mean ages were 23.8 years old (SD = 4.0 years), 25.0 years old (SD = 5.7 years), 22.5 years old (SD = 4.7 years), and 27.8 years old (SD = 6.3 years), respectively.

To abide by the social-distancing law during the COVID-19 pandemic, data collection switched from face-to-face to an online experimental platform (*Gorilla*; Anwyl-Irvine *et al.*, 2020). An automatic procedure ensured that the participants could only access the platform with a desktop or laptop computer. Prior to the experiment, the participants were asked to stay in a quiet environment and calibrate their headphones/earphones to a comfortable volume. To monitor their attention, we pseudo-randomly inserted five attention-check trials throughout the experiment. On each trial, participants indicated whether the two sounds they heard were the same. The two sounds were always acoustically identical, so the attention-check trials could be answered accurately with adequate attention. To be empirically stringent, we had excluded participants who made more than one mistake, i.e., below 80% accuracy. Five Cantonese non-musicians and three English musicians had been excluded from the dataset for this reason. The mean attention-check accuracy was very high in the remaining Cantonese musicians [mean (M) = 97.33%, SD = 6.92%], Cantonese non-musicians (M = 96.80%, SD = 7.48%), English musicians (M = 94.07%, SD = 9.31%), and English non-musicians (M = 96.00%, SD = 8.14%).

B. Modified Thai tone AX discrimination task

1. Stimuli

Two native Thai speakers (one male and one female) recorded the stimuli in a soundproof room at The University of Hong Kong. They naturally produced the five Thai tones (mid, low, falling, high, and rising) in each of the four Thai syllables (/tɕʰɔː/, /kʰɯː/, /tʰiːa/, and /wuːa/). There were altogether 40 stimuli (5 tones × 4 syllables × 2 speakers). The acoustic profiles of these stimuli were described in the previous study (Choi and Tsui, 2022).

2. Stimuli presentation

The task comprised a control block and an orthogonal block. On each trial, listeners heard two audio stimuli with an inter-stimulus interval of 500 ms. After the second stimulus was fully presented, the listeners were asked to judge as quickly as possible whether the two audio stimuli were identical or different. To prevent the use of an *ad hoc* acoustic strategy, the two audio stimuli on each trial were always produced by different speakers. Even on the "identical" trials, the two audio stimuli were acoustically distinct. Across the trials, the gender order (male-female or female-male) was counterbalanced. In the control block, the segmental

information was constant on each trial, e.g., /k^hɯː-high/–/k^hɯː-low/ or /t^hiːa-high/–/t^hiːa-low/ for a "different" trial and /k^hɯː-high/–/k^hɯː-high/ or /t^hiːa-high/–/t^hiːa-high/ for an "identical" trial.

In the orthogonal block, each trial contained two stimuli with different segmental information. Their tones could be identical or different, e.g., /k^hɯː-high/–/ t^hiːa-low/ for a "different" trial and /k^hɯː-high/–/ t^hiːa-high/ for an "identical" trial. Listeners were asked to ignore the segmental variations and judge only the tones. In the control and orthogonal blocks, the frequencies of occurrence of each syllable (/tɕ^{h}ɔː/, /k^hɯː/, /t^hiːa/, and /wuːa/) and Thai tone contrast (mid-low, mid-falling, mid-high, mid-rising, low-falling, low-high, low-rising, falling-high, falling-rising, and high-rising) were equal. Each block began with three practice trials and 80 experimental trials (5 tones × 4 syllables × 2 speakers × 2 repetitions). The order of presentation of the two blocks was counterbalanced across participants. On each experimental trial, the accuracy and response time were recorded. The internal consistencies were high in the control (Cronbach's $\alpha = 0.86$) and orthogonal blocks (Cronbach's $\alpha = 0.80$).

III. RESULTS

A. Computation of *d′*

In each block of the modified Thai tone AX discrimination task, the d' was calculated by subtracting the z-transform of the hit rate by that of the false alarm rate (MacMillan and Creelman, 2005). Each correct "different" trial was a hit, whereas each incorrect "identical" trial was a false alarm. To avoid infinity d', we manually adjusted perfect hit (1.0) and false alarm (0) rates to 0.995 and 0.005, respectively (Bidelman *et al.*, 2013; Choi *et al.*, 2019). Figure 1 summarizes the mean d' and response time of the groups in the control and orthogonal blocks.

B. Conventional analyses of *d′* and response time across blocks

To examine how language experience and musicianship shape tone-segmental perceptual integrality, we conducted a three-way mixed analysis of variance (ANOVA) on d' with block (control and orthogonal) as the within-subject factor and language (Cantonese and English) and musicianship (musicians and non-musicians) as the between-subjects factors (see Fig. 1, top). ANOVA revealed significant main effects of block, $F(1, 108) = 43.91$, $p < 0.001$, $\eta_p^2 = 0.29$; language, $F(1, 108) = 32.03$, $p < 0.001$, $\eta_p^2 = 0.23$; and musicianship, $F(1, 108) = 8.48$, $p = 0.004$, $\eta_p^2 = 0.07$. The interaction between block and language was significant, $F(1, 108) = 48.71$, $p < 0.001$, $\eta_p^2 = 0.31$, but not the interaction between block and musicianship, $p = 0.589$; language and musicianship, $p = 0.466$; and the three-way interaction, $p = 0.290$.

We conducted simple effect analysis to unpack the interaction between block and language. The Cantonese listeners showed a significantly lower d' in the orthogonal block than in the control block, $p < 0.001$. However, the English listeners showed similar d' values across the orthogonal and control blocks, $p = 0.801$. These findings indicate that the Cantonese listeners perceived tones and segmental information integrally, whereas the English listeners perceived them independently. Additionally, the main effect of

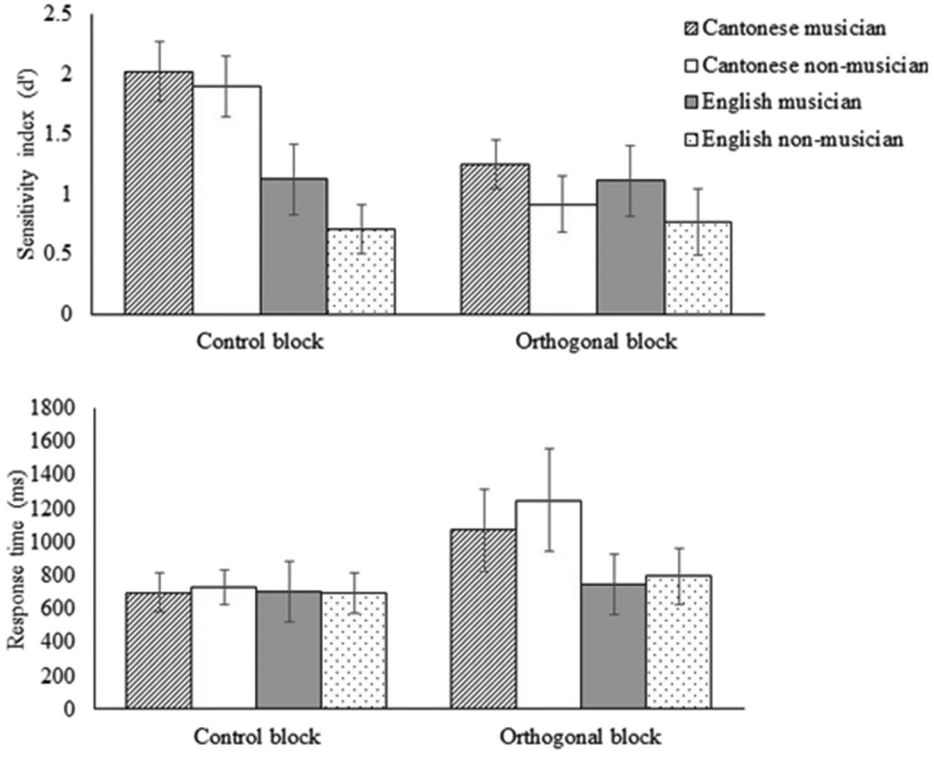

FIG. 1. Mean d' (top), and response time (bottom) of the Cantonese and English listeners. Error bars, 95% confidence interval (CI).

https://doi.org/10.1121/10.0020579

musicianship indicated that the musicians outperformed the non-musicians across the two blocks.

To supplement the above analysis, we further conducted three-way mixed ANOVAs on response time with block (control and orthogonal) as the within-subject factor and language (Cantonese and English) and musicianship (musicians and non-musicians) as the between-subjects factors (see Fig. 1, bottom). There were significant main effects of block, $F(1, 108) = 30.78$, $p < 0.001$, $\eta_p^2 = 0.22$, and language, $F(1, 108) = 6.70$, $p = 0.011$, $\eta_p^2 = 0.06$. The interaction between block and language was significant, $F(1, 108) = 16.02$, $p < 0.001$, $\eta_p^2 = 0.13$. However, the main effect of musicianship, $p = 0.423$; the interaction between language and musicianship, $p = 0.592$; and the three-way interaction, $p = 0.609$, were not significant.

We conducted simple effect analysis to unpack the interaction between block and language. Relative to their own performance in the control block, the Cantonese listeners were significantly slower in discriminating Thai tones in the orthogonal block, $p < 0.001$. However, the English listeners showed similar response time across the orthogonal and control blocks, $p = 0.272$. Consistent with the d' analysis, the Cantonese listeners perceived tones and segmental information integrally, whereas the English perceived them independently.

C. Garner *d'* and RT indexes

To avoid overcomplicating the Bayesian models (see Sec. III F), we devised the Garner d' and RT indexes to quantify the Garner effects on d' and response time. For each participant, the Garner d' index was obtained by subtracting the d' in the orthogonal block from that in the control block. The more positive the value of the Garner d' index, the larger the Garner effect is on the participant's d'. Similarly, the Garner RT index was calculated by subtracting the response time in the orthogonal block from that in the control block. The more negative the value of the Garner RT index, the larger the Garner effect is on the participant's response time. Figures 2 and 3 summarize the mean Garner d' and RT indexes of the four groups.

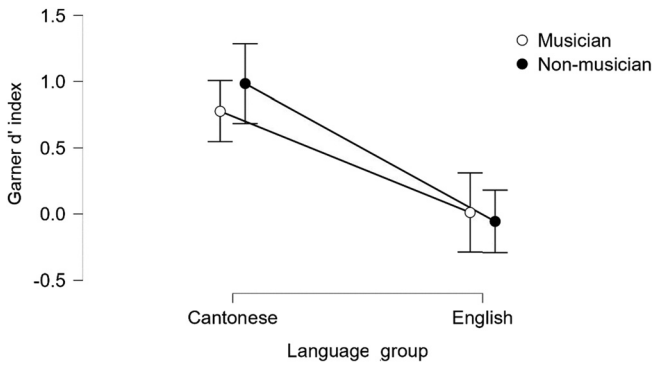

FIG. 2. Mean Garner d' index of the Cantonese and English listeners. Error bars, 95% CI.

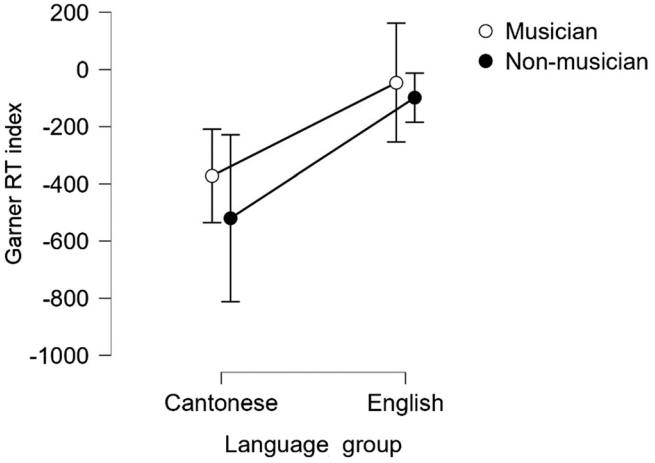

FIG. 3. Mean Garner RT index of the Cantonese and English listeners. Error bars, 95% CI.

D. Conventional analysis of Garner *d'* and RT indexes

We conducted a two-way mixed ANOVA on Garner d' index with language (Cantonese and English) and musicianship (musicians and non-musicians) as the between-subjects factors. The main effect of language was significant, $F(1, 108) = 48.71$, $p < 0.001$, $\eta_p^2 = 0.31$, but not the main effect of musicianship, $p = 0.589$, and the interaction between language and musicianship, $p = 0.290$. Consistent with the previous study, the Cantonese listeners experienced a larger Garner interference effect than the English listeners, $p < 0.001$ (Choi and Tsui, 2022). Pertinent to our research questions, musicianship had no effect on tone-segmental perceptual integrality in Cantonese and English listeners.

To supplement the above analysis, we conducted the same set of ANOVAs on Garner RT index. The main effect of language was significant, $F(1, 108) = 16.02$, $p < 0.001$, $\eta_p^2 = 0.13$, but not the main effect of musicianship, $p = 0.286$, and the interaction between language and musicianship, $p = 0.609$. Consistent with the Garner d' index analysis, musicianship did not significantly influence tone-segmental perceptual integrality in Cantonese and English listeners. These results are fully consistent with the three-way ANOVAs reported in Sec. III B.

E. Analytic issues and statistic strategies

The research question is whether musicianship influences tone-segmental perceptual integrality in Cantonese and English listeners. From a statistical perspective, the question is whether there is a main effect of musicianship and/or its interaction with language on the Garner indexes. If the main effect of musicianship (or interaction thereof) is present, one can conclude that musicianship influences tone-segmental perceptual integrality in Cantonese and English listeners. If otherwise, one can conclude that musicianship does not influence their perceptual integrality. In Sec. III D, we reported the data analysis based on the null-hypothesis significance testing that has been conventionally

TABLE I. Bayes factor and evidence strength.

Bayes factor (BF_{10})	Hypothesis supported	Evidence strength
>100	Alternative hypothesis	Extreme
30–100	Alternative hypothesis	Very strong
10–30	Alternative hypothesis	Strong
3–10	Alternative hypothesis	Moderate
1–3	Alternative hypothesis	Anecdotal
1	None	No evidence
1/3–1	Null hypothesis	Anecdotal
1/10–1/3	Null hypothesis	Moderate
1/30–1/10	Null hypothesis	Strong
1/100–1/30	Null hypothesis	Very strong
<1/100	Null hypothesis	Extreme

used for psycholinguistic research. The limitation of null-hypothesis significance testing is that it can only reject or not reject a null hypothesis (Ly *et al.*, 2019). Although the *p*-value is larger than 0.05, researchers cannot conclude the absence of an effect. To overcome this challenge, we have adopted Bayesian hypothesis testing and will report it below (Wagenmakers, 2007). Unlike null hypothesis significance testing, the Bayesian approach quantifies evidence both for alternative and null hypotheses (Dienes, 2014, 2016). In other words, it can provide not only evidence of presence but also evidence of absence of the main effect of musicianship and its interaction with language.

Bayes factor (BF_{10}) refers to the ratio of predictive adequacies of two competing statistical models (Schönbrodt and Wagenmakers, 2018). A BF_{10} of 10 indicates that the data are ten times more likely under the alternative hypothesis than under the null hypothesis. Conversely, a BF_{10} of 0.1 indicates that the data are ten times more likely under the null hypothesis than the alternative hypothesis. Lee and Wagenmakers (2013) heuristically classified different values of BF_{10} as different levels of evidence (see Table I).

F. Bayesian analysis of Garner *d′* index

To address the research question, we conducted Bayesian two-way between-subjects ANOVA with JASP 0.17.1 (Love *et al.*, 2019). The dependent variable was Garner d' index, and the fixed factors were language (Cantonese and English) and musicianship (musicians and non-musicians). There were four models in total. Model 1 only includes the main effect of language. Model 2 includes the main effects of language and musicianship. Model 3 includes the two main effects and the interaction between language and musicianship. Model 4 only includes the main effect of musicianship. Given the lack of previous findings, we adopted a uniform prior that is relatively objective (Jaynes, 2003). Below, we summarize the results of the model comparison relative to the null hypothesis and the best-fit model. Annotated .jasp files, including distribution plots, data, and input options, are available on Open Science Framework (https://osf.io/zqdrj/?view_only=5bd911ccddd24fcfb02e853f4babcd63).

Table II shows the model comparison relative to the null hypothesis. Bayes factors indicate that the data best support model 1 ($B_{10} = 3.07 \times 10^7$), followed by model 2 ($B_{10} = 7.22 \times 10^6$) and model 3 ($BF_{10} = 2.97 \times 10^6$), each of which has extreme evidence for the alternative hypothesis. The data do not support model 4 ($BF_{10} = 0.20$), with moderate evidence for the null hypothesis.

Table III shows the model comparison relative to the best-fit model (model 1). Bayes factors indicate that the data are less likely observed under model 2 with moderate evidence ($BF_{10} = 0.24$), under model 3 with strong evidence ($BF_{10} = 0.097$), and under model 4 ($BF_{10} = 6.52 \times 10^{-9}$) and null model ($BF_{10} = 3.25 \times 10^{-8}$) with extreme evidence. As shown in Table IV, the analysis of effects yields extreme evidence for the main effect of language ($BF_{incl} = 2.27 \times 10^7$) and moderate evidence against the main effect of musicianship ($BF_{incl} = 0.22$) and its interaction with language ($BF_{incl} = 0.31$). Taken together, Bayesian analysis reveals (i) the presence of the main effect of language and (ii) the absence of main effect of musicianship and its interaction with language.

G. Bayesian analysis of Garner RT index

To supplement the above results, we conducted the same set of analysis on Garner RT index. Table V shows the model comparison relative to the null hypothesis. Bayes factors indicate that the data best support model 1 ($BF_{10} = 157.36$) with extreme evidence for the alternative hypothesis. The second- and third-best-fit models are models 2 ($BF_{10} = 51.57$) and 3 ($BF_{10} = 14.62$) with very strong and strong evidence, respectively. The data do not support

TABLE II. Model comparison relative to null model of Garner d' index.

Models	P(M)	P(M|data)	BF_M	BF_{10}	Error (%)
Null model	0.20	2.44×10^{-8}	9.77×10^{-8}	1.00	
Model 1 (LG)[a]	0.20	0.75	12.06	3.07×10^7	1.97×10^{-10}
Model 2 (LG + MS)[b]	0.20	0.18	0.86	7.22×10^6	4.48
Model 3 (LG + MS + LG*MS)	0.20	0.07	0.31	2.97×10^6	1.32
Model 4 (MS)	0.20	4.90×10^{-9}	1.96×10^{-8}	0.20	0.03

[a]Language (LG).
[b]Musicianship (MS).

https://doi.org/10.1121/10.0020579

TABLE III. Model comparison relative to best-fit model of Garner d' index.

Models	P(M)	P(M|data)	BF_M	BF_{10}	Error (%)
Model 1 (LG)	0.20	0.75	12.06	1.00	
Model 2 (LG + MS)	0.20	0.18	0.86	0.24	1.00
Model 3 (LG + MS + LG*MS)	0.20	0.07	0.31	0.097	1.32
Null model	0.20	2.44×10^{-8}	9.77×10^{-8}	3.25×10^{-8}	1.97×10^{-10}
Model 4 (MS)	0.20	4.90×10^{-9}	1.96×10^{-8}	6.52×10^{-9}	0.03

model 4 ($BF_{10} = 0.26$) with moderate evidence for the null hypothesis.

Table VI shows the model comparison relative to the best-fit model (model 1). Bayes factors indicate that the data are less likely observed under model 2 ($BF_{10} = 0.328$) with moderate evidence, under model 3 ($BF_{10} = 0.09$) with strong evidence, and under model 4 ($BF_{10} = 0.002$) and null model ($BF_{10} = 0.01$) with extreme evidence. As shown in Table VII, the analysis of effects yields extreme evidence for the main effect of language ($BF_{incl} = 118.75$) and moderate evidence against the main effect of musicianship ($BF_{incl} = 0.28$) and its interaction with language ($BF_{incl} = 0.28$). Consistent with the Garner d' index analysis, Bayesian analysis on Garner RT index reveals the main effect of language and the absence of main effect of musicianship and its interaction with language.

IV. DISCUSSION

The present study investigated whether musicianship influenced non-native tone-perceptual integrality in Cantonese and English listeners. There were three possibilities, i.e., musicianship enhances, inhibits, or does not affect perceptual integrality. As detailed below, our findings support the third possibility.

The principal finding is that musicianship does not influence tone-segmental perceptual integrality. Consistent with previous research, the Cantonese listeners integrally perceived Thai tones and segmental information, whereas the English listeners showed independent perception (Choi and Tsui, 2022). We originally believed that musicianship might enhance (or inhibit) tone-segmental perceptual integrality in Cantonese listeners and English listeners. Contrary to our expectation, the Cantonese musicians and Cantonese non-musicians showed similar degrees of tone-segmental perceptual integrality. Furthermore, both English musicians and English non-musicians perceived Thai tones and segmental information independently. Collectively, musicianship neither enhanced nor decreased tone-segmental perceptual integrality. In other words, musicianship simply has no effect on tone-segmental perceptual integrality.

TABLE IV. Analysis of effects in Garner d' index.

Effects	P(incl)	P(excl)	P(incl|data)	P(excl|data)	BF_{incl}
LG	0.60	0.40	1.00	2.93×10^{-8}	2.27×10^{7}
MS	0.60	0.40	0.25	0.75	0.22
LG*MS	0.20	0.80	0.07	0.93	0.31

Based on past and present findings, the effect of musicianship on perceptual integrality seems to be restricted to the music domain. In a previous MMN additivity study, Dutch musicians and non-musicians completed two passive oddball paradigms, one with musically relevant stimuli (using Alberti bass as standard; see Hansen *et al.*, 2022, p. 4) and one with musically less-relevant stimuli (using chromatic notes as standard). In the musically relevant paradigm, the musicians showed more integral perception of music features than the non-musicians (Hansen *et al.*, 2022). However, in the musically less-relevant paradigm, the musicians no longer showed more integral perception than the non-musicians. Extending the previous study, our findings suggest that the musicianship effect is not transferrable to the language domain. Instead, the effect of musicianship on perceptual integrality only applies to domain-relevant stimuli.

From a theoretical perspective, there is little need to introduce a music element into the dimensional transfer hypothesis (Choi and Tsui, 2022). In the analyses of Garner d' and RT indexes, the main effect of language revealed that the Cantonese listeners had a higher degree of tone-segmental perceptual integrality than the English listeners. Consistent with the previous study, the present results indicate that first language experience shapes tone-segmental perceptual integrality in non-native speech (Choi and Tsui, 2022). Extending the previous study, our results further added that musicianship does not seem to play a significant role, if any. In its current form, the dimension transfer hypothesis posits that first language functional importance determines whether foreign suprasegmental and segmental information are integrally or independently perceived (Choi and Tsui, 2022). In Cantonese, segmental and tonal information are important constituents of phonological representations (Shen *et al.*, 2021). To maximally support word form detection, Cantonese listeners are perceptually wired to integrate segmental and tonal information. English has no tones, and only very few words are suprasegmentally contrastive

TABLE V. Model comparison relative to null model of Garner RT index.

Models	P(M)	P(M|data)	BF_M	BF_{10}	Error (%)
Null model	0.20	0.004	0.02	1.00	
Model 1 (LG)	0.20	0.70	9.33	157.36	1.14×10^{-8}
Model 2 (LG + MS)	0.20	0.23	1.19	51.57	0.83
Model 3 (LG + MS + LG*MS)	0.20	0.07	0.28	14.62	2.40
Model 4 (MS)	0.20	0.001	0.01	0.26	0.03

TABLE VI. Model comparison relative to best-fit model of Garner RT index.

Models	P(M)	P(M|data)	BF_M	BF_{10}	Error (%)
Model 1 (LG)	0.20	0.70	9.33	1.000	
Model 2 (LG + MS)	0.20	0.23	1.19	0.328	0.83
Model 3 (LG + MS + LG*MS)	0.20	0.07	0.28	0.09	2.40
Null model	0.20	0.004	0.02	0.01	1.14×10^{-8}
Model 4 (MS)	0.20	0.001	0.01	0.002	0.03

(e.g., noun and verb forms of *import* and *suspect*) (Cutler, 2012). Thus, our English listeners processed segmental and tonal information independently as separate information. Our study further shows that musicianship does not alter how they integrate (or not) segmental and tonal information. In the theoretical aspect, the present findings suggest that the dimension transfer hypothesis has adequately accounted for tone-segmental perceptual integrality. As such, there is little need to incorporate the musicianship element into the hypothesis.

In addition to the dimension transfer hypothesis, other theories may account for the present findings too. According to the native language neural commitment (NLNC) theory, early language experience wires a listener's neural circuitry for optimal perception and acquisition of first language phonological structures (Kuhl, 2010; Zhang *et al.*, 2005). For our Cantonese listeners, they have a neural commitment to process Cantonese tones and segmental information integrally (Choi *et al.*, 2017; Yu *et al.*, 2022). With extensive Cantonese experience, the Cantonese listeners may have developed strong neural representations of tonal information that are difficult to ignore even when attending to other dimensions of the stimuli. Crucially, such a neural commitment is transferrable to non-native speech perception. Like the dimensional transfer hypothesis, NLNC theory posits that first language neural commitment is malleable, for example, by second language learning (Zhang *et al.*, 2009; Zhang *et al.*, 2011). Indeed, Dutch listeners with extensive Mandarin learning experience perceived Mandarin tones and segmental information integrally, whereas Mandarin-naive Dutch listeners did not (Zou *et al.*, 2017). Our behavioral results further imply that musicianship does not alter the neural commitment on tone-segmental integrality.

Future studies can seek to replicate the current findings in the opposite direction. Previous studies showed Garner interference in tone and segmental perception tasks (e.g., Repp and Lin, 1990; Tong *et al.*, 2008). In the present study, we only tested the effect of task-irrelevant segmental variation on tone discrimination. To supplement the current evidence, future studies can test Cantonese and English listeners in the opposite direction, i.e., requesting them to discriminate Thai segmental information while ignoring task-irrelevant tonal variations. Based on the current findings, we expect that musicianship would not alter the Garner effect (or the lack thereof) of tone variation on segmental discrimination. Due to the COVID pandemic, we could only test participants online with *post hoc* attention monitoring. Whenever possible, future studies should test participants face-to-face in a well-controlled acoustic environment.

TABLE VII. Analysis of effects in Garner RT index.

Effects	P(incl)	P(excl)	P(incl|data)	P(excl|data)	BF_{incl}
LG	0.60	0.40	0.99	0.01	118.75
MS	0.60	0.40	0.30	0.70	0.28
LG*MS	0.20	0.80	0.07	0.94	0.28

In conclusion, musicianship has no effect on tone-segmental perceptual integrality in Cantonese and English listeners. Consistent with previous research, tone-segmental perceptual integrality hinges on first language experience (Choi and Tsui, 2022; Lin and Francis, 2014). While musicianship (Choi, 2020; Delogu *et al.*, 2010; Zheng and Samuel, 2018) or music training (Moreno *et al.*, 2009; Nan *et al.*, 2018) apparently enhances speech sensitivity, not all musical advantages are transferrable to the language domain. From the theoretical perspective, the dimensional transfer hypothesis has adequately accounted for tone-segmental integrity in non-native speech perception (Choi and Tsui, 2022). There is little need to add a musicianship element to the hypothesis.

ACKNOWLEDGMENTS

We would like to acknowledge Shirlyn Chiu, Sze Hung Chung, Kin Chak Lau, Anson Chun Fai Leung, and Mei Sze Wong for their assistance in data collection and entry. Publication was made possible in part by support from the HKU Libraries Open Access Author Fund (OAAF2023-24-012) sponsored by the HKU Libraries, Seed Fund for Basic Research for New Staff (202107185043) sponsored by The University of Hong Kong, and Project-based Research Funding sponsored by the Faculty of Education of The University of Hong Kong. The authors declare no conflicts of interest. This study was approved by the University Human Research Ethics Committee. The data that support the findings of this study are openly available in the Open Science Framework, reference number 5bd911ccddd24fcfb02e853f4babcd63.

Ahissar, M., and Hochstein, S. (**2004**). "The reverse hierarchy theory of visual perceptual learning," Trends Cognit. Sci. **8**(10), 457–464.

Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (**2009**). "Reverse hierarchies and sensory learning," Philos. Trans. R. Soc.: Biol. Sci. **364**(1515), 285–299.

Alexander, J. A., Wong, P. C. M., and Bradlow, A. R. (**2005**). "Lexical tone perception in musicians and non-musicians," in *Proceedings of the 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal (ISCA, Lisbon).

Algom, D., and Fitousi, D. (**2016**). "Half a century of research on Garner interference and the separability–integrality distinction," Psychol. Bull. **142**(12), 1352–1383.

Allen, E. J., and Oxenham, A. J. (**2014**). "Symmetric interactions and interference between pitch and timbre," J. Acoust. Soc. Am. **135**(5), 1371–1379.

Anwyl-Irvine, A. L., Massonnie, J., Flitton, A., Kirkham, N., and Evershed, J. K. (**2020**). "Gorilla in our midst: An online behavioral experiment builder," Behav. Res. Methods **52**(1), 388–407.

Berkowitz, S., Fontrier, G., and Kraft, L. (**1997**). *A New Approach to Sightsinging*, 4th ed. (W. W. Norton & Company, New York).
Bidelman, G. M., Hutka, S., and Moreno, S. (**2013**). "Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music," PLoS One **8**, e60676.
Choi, W. (**2020**). "The selectivity of musical advantage: Musicians exhibit perceptual advantage for some but not all Cantonese tones," Music Percept. **37**(5), 423–434.
Choi, W. (**2021**). "Musicianship influences language effect on musical pitch perception," Front. Psychol. **12**, 712753.
Choi, W. (**2022a**). "Towards a native OPERA hypothesis: Musicianship and English stress perception," Lang. Speech **65**(3), 697–712.
Choi, W. (**2022b**). "What is 'music' in music-to-language transfer? Musical ability but not musicianship supports Cantonese listeners' English stress perception," J. Speech Lang. Hear. Res. **65**, 4047–4059.
Choi, W., Tong, X., Gu, F., Tong, X., and Wong, L. (**2017**). "On the early neural perceptual integrality of tones and vowels," J. Neurolinguistics **41**, 11–23.
Choi, W., Tong, X., and Samuel, A. G. (**2019**). "Better than native: Tone language experience enhances English lexical stress discrimination in Cantonese-English bilingual listeners," Cognition **189**, 188–192.
Choi, W., and Tsui, K. Y. R. (**2022**). "Perceptual integrality of foreign segmental and tonal information: Dimensional transfer hypothesis," Stud. Second Lang. Acquis. (published online).
Cooper, A., and Wang, Y. (**2012**). "The influence of linguistic and musical experience on Cantonese word learning," J. Acoust. Soc. Am. **131**(6), 4756–4769.
Cooper, A., Wang, Y., and Ashley, R. (**2017**). "Thai rate-varied vowel length perception and the impact of musical experience," Lang. Speech **60**(1), 65–84.
Cutler, A. (**2012**). *Native Listening: Language Experience and the Recognition of Spoken Words* (MIT, Cambridge, MA).
Delogu, F., Lampis, G., and Belardinelli, M. (**2010**). "From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception," Eur. J. Cogn. Psychol. **22**(1), 46–61.
Dienes, Z. (**2014**). "Using Bayes to get the most out of nonsignificant results," Front. Psychol. **5**, 781.
Dienes, Z. (**2016**). "How Bayes factors change scientific practice," J. Math. Psychol. **72**, 78–89.
Freeman, V. (**2023**). "Production and perception of prevelar merger: Two-dimensional comparisons using Pillai scores and confusion matrices," J. Phonetics **97**, 101213.
Gao, S., Hu, J., Gong, Q., Chen, S., Kendrick, K. M., and Yao, D. (**2012**). "Integration of consonant and pitch processing as by the absence of additivity in mismatch negativity," PLoS One **7**(5), e38289.
Garner, W. R. (**1974**). *The Processing of Information and Structure* (Wiley, New York).
Hansen, N. C., Højlund, A., Møller, C., Pearce, M., and Vuust, P. (**2022**). "Musicians show more integrated neural processing of contextually relevant acoustic features," Front. Neurosci. **16**, 907540.
Jaynes, E. T. (**2003**). *Probability Theory: The Logic of Science* (Cambridge University, Cambridge, UK).
Kolinsky, R., Cuvelier, H., Goetry, V., Peretz, I., and Morais, J. (**2009**). "Music training facilitates lexical stress processing," Music Percept. **26**(3), 235–246.
Kuhl, P. K. (**2010**). "Brain mechanisms in early language acquisition," Neuron **67**(5), 713–727.
Lee, L., and Nusbaum, H. C. (**1993**). "Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese," Percept. Psychophys. **53**(2), 157–165.
Lee, M. D., and Wagenmakers, E. J. (**2013**). *Bayesian Cognitive Modeling: A Practical Course* (Cambridge University, Cambridge, UK).
Lin, M., and Francis, A. L. (**2014**). "Effects of language experience and expectations on attention to consonants and tones in English and Mandarin Chinese," J. Acoust. Soc. Am. **136**(5), 2827–2838.
Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., and Wagenmakers, E. J. (**2019**). "JASP: Graphical statistical software for common statistical designs," J. Stat. Softw. **88**(2), 1–17.
Ly, A., Etz, A., Marsman, M., and Wagenmakers, E. (**2019**). "Replication Bayes factors from evidence updating," Behav. Res. **51**, 2498–2508.
Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, 2nd ed. (Psychological Press, New York).
McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception.,"Cognit. Psychol. **18**(1), 1–86.
Melara, R., and Marks, L. E. (**1990**). "Interaction among auditory dimensions: Timbre, pitch, and loudness," Percept. Psychophys. **48**, 169–178.
Miya, F. M. (**2005**). "Educational content in Kenyan performing arts: The case of *Isukuti* performing art," in *Proceedings of the East African Symposium on Music Education*, Nairobi, Kenya (Kenyatta University, Nairobi), pp. 80–94.
Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (**2009**). "Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity," Cereb. Cortex **19**(3), 712–723.
Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., Gabrieli, J. D. E., and Desimone, R. (**2018**). "Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children," Proc. Natl. Acad. Sci. U.S.A. **115**(28), E6630–E6639.
Neuhaus, C., and Knösche, T. R. (**2008**). "Processing of pitch and time sequences in music," Neurosci. Lett. **441**(1), 11–15.
Ottman, R. W. (**1986**). *Music for Sight-Singing*, 3rd ed. (Prentice-Hall, Englewood Cliffs, NJ).
Patel, A. D. (**2011**). "Why would musical training benefit the neural encoding of speech? The OPERA hypothesis," Front. Psychol. **2**, 142.
Patel, A. D. (**2014**). "Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis," Hear. Res. **308**, 98–108.
Pitt, M. (**1994**). "Perception of pitch and timbre by musically trained and untrained listeners," J. Exp. Psychol.: Hum. Percept. Perform. **20**, 976–986.
Prince, J. B. (**2011**). "The integration of stimulus dimensions in the perception of music," Q. J. Exp. Psychol. **64**(11), 2125–2152.
Repp, B. H., and Lin, H. B. (**1990**). "Integration of segmental and tonal information in speech perception: A cross-linguistic study," J. Phon. **18**(4), 481–495.
Sadakata, M., and Sekiyama, K. (**2011**). "Enhanced perception of various linguistic features by musicians: A cross-linguistic study," Acta Psychol. **138**(1), 1–10.
Schönbrodt, F. D., and Wagenmakers, E. J. (**2018**). "Bayes factor design analysis: Planning for compelling evidence," Psychon. Bull. Rev. **25**, 128–142.
Shen, W., Hyönä, J., Wang, Y., Hou, M., and Zhao, J. (**2021**). "The role of tonal information during spoken-word recognition in Chinese: Evidence from a printed-word eye-tracking study," Mem. Cogn. **49**(1), 181–192.
Sinagra, C., and Wiener, S. (**2022**). "The perception of intonational and emotional speech prosody produced with and without a face mask: an exploratory individual differences study," Cognit. Res.: Princ. Implic. **7**, 89.
Tierney, A. T., and Kraus, N. (**2014**). "Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH)," Front. Hum. Neurosci. **8**(8), 949.
Tong, Y., Francis, A. L., and Gandour, J. T. (**2008**). "Processing dependencies between segmental and suprasegmental features in Mandarin Chinese," Lang. Cogn. **23**(5), 689–708.
Wagenmakers, E. J. (**2007**). "A practical solution to the pervasive problems of *p* values," Psychon. Bull. Rev. **14**(5), 779–804.
Yu, K., Chen, Y., Wang, M., Wang, R., and Li, L. (**2022**). "Distinct but integrated processing of lexical tones, vowels, and consonants in tonal language speech perception: Evidence from mismatch negativity," J. Neurolinguistics **61**, 101039.
Yu, L., and Zhang, Y. (**2018**). "Testing native language neural commitment at the brainstem level: A cross-linguistic investigation of the association between frequency-following response and speech perception," Neuropsychologia **109**, 140–148.
Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., and Carney, E. (**2011**). "Neural coding of formant-exaggerated speech in the infant brain," Dev. Sci. **14**(3), 566–581.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., and Nemoto, I. (**2009**). "Neural signatures of phonetic learning in adulthood: A magnetoencephalography study," NeuroImage **46**(1), 226–240.
Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., and Tohkura, Y. (**2005**). "Effects of language experience: Neural commitment to language-specific auditory patterns," NeuroImage **26**(3), 703–720.
Zheng, Y., and Samuel, A. G. (**2018**). "The effects of ethnicity, musicianship, and tone language experience on pitch perception," Q. J. Exp. Psychol. **71**(12), 2627–2642.
Zou, T., Chen, Y., and Caspers, J. (**2017**). "The developmental trajectories of attention distribution and segment-tone integration in Dutch learners of Mandarin tones," Bilingualism **20**(5), 1017–1029.