

RESEARCH REPORT  

Perceptual integrality of foreign segmental and tonal information: Dimensional transfer hypothesis

William Choi^{1*}  and Rachel Ka-Ying Tsui² 

¹Academic Unit of Human Communication, Development, and Information Sciences, the University of Hong Kong, Pokfulam, Hong Kong; ²Laboratory for Language Development, RIKEN Center for Brain Science, Saitama, Japan

*Corresponding author. E-mail: wilchoi@hku.hk

(Received 13 February 2022; Revised 25 October 2022; Accepted 21 November 2022)

Abstract

This study investigates whether (a) Cantonese and (b) English listeners integrally or independently perceive Thai tone and segmental information. Listeners completed a modified AX discrimination task that contained a control block (without segmental variation) and an orthogonal block (with segmental variation). Relative to their own performance in the control block, the Cantonese listeners showed increased response time, decreased proportion of accuracy, and decreased sensitivity index in the orthogonal block. By contrast, the English listeners showed similar response time, proportion of accuracy, and sensitivity index across the two blocks. These reflect integral processing among the Cantonese but not the English listeners. This finding motivates the dimensional transfer hypothesis. The hypothesis posits that L1 perceptual experience shapes the perceptual integrality (or nonintegrality) of foreign suprasegmental and segmental information.

Introduction

Nonnative speech perception theories largely conceptualize cross-linguistic transfer on a microscopic scale, for example, how nonnative phonemes and tones are mapped into L1 phonological categories (perceptual assimilation model, Best, 1995; perceptual assimilation model for suprasegmentals, So & Best, 2010, 2014), and how L1 experience shapes listeners' choice of acoustic cue for nonnative speech perception (cue-weighting transfer hypothesis, Chang, 2018; Chrabaszcz et al., 2014; Kim & Tremblay, 2021; acoustic-attentional-contextual hypothesis, Choi, 2022a). Going beyond individual phonemes and acoustic cues, the macroscopic scale concerns how segmental and suprasegmental information are processed. For instance, tone language listeners process L1 segmental and suprasegmental information integrally (Choi et al., 2017; Gao et al., 2012; Repp & Lin, 1990). Regarding nonnative speech perception, it remains unclear whether such perceptual integrality is transferrable across languages. Filling

this research gap provides grist for theorizing cross-linguistic transfer both microscopically and macroscopically. Thus, we investigate whether Cantonese and English listeners integrally or independently process tonal and segmental information in Thai—a language unknown to both groups of listeners. Methodologically, the innovative part of the study lies in our modified AX task. As described later, our task gives a more sensitive measure of perceptual integrality than the conventional speeded classification task (Garner, 1974).

Source-filter theory

From the macroscopic perspective, the source-filter theory represents speech production as a combination of two highly independent mechanisms—phonation and articulation (Ladefoged & Johnson, 2011; Tokuda, 2021; see also Chiba & Kajiyama, 1942). At the laryngeal level, the vibration of vocal folds generates an undifferentiated sound stream (*source*). The vibration rate of the vocal folds determines the fundamental frequency (f_0) of the sound stream. In tone languages such as Cantonese, variations of f_0 (or *tone*) of a syllable can result in different meanings, for example, /fu1/ 夫 *husband*, /fu2/ 苦 *bitter*, /fu3/ 褲 *trousers*, /fu4/ 扶 *support*, /fu5/ 婦 *woman*, and /fu6/ 父 *father* (Yip, 2002). At the supralaryngeal level, articulators constrict the airstream and filter the source spectrum (*filter*). The differentiation of sound stream gives rise to different consonants and vowels, that is, segmental information. Segmental and suprasegmental (tone) information are important phonological constituents of Cantonese words (Yip, 2002). Consistent with this notion, eye-tracking and lexical decision studies have shown that Cantonese listeners attend to both source and filter for spoken word recognition (Ho, 2015; Shen et al., 2021). This raises the question of whether source and filter are integrally processed.

Behavioral and neural evidence of perceptual integration

Perceptual integration arises from the classic *combination problem* that concerns how multiple features are combined into a single perceptual entity (Bayne & Chalmers, 2003; Revonsuo & Newman, 1999). The combination problem has captured much attention in the visual perception field, epitomized by the feature integration theory (Treisman & Gelade, 1980). The theory argues that visual features (e.g., tail, legs, and fur) are independently processed as separate streams at the preattentive stage. At the attentive stage, the features are combined into a single percept (e.g., dog), that is, integration. Analogous to the visual features, tone language listeners must combine source (tone) and filter (consonant and vowels) to recognize a spoken word (Ho, 2015; Shen et al., 2021). Does this integration occur at the preattentive or attentive level?

Behavioral studies on perceptual integration have utilized the Garner paradigm (Lee & Nusbaum, 1993; Lin & Francis, 2014; Repp & Lin, 1990; Tong et al., 2008) or its variation (Zou et al., 2017). Typically, listeners complete both control and orthogonal blocks of the speeded classification or AXB discrimination task. In the speeded classification task, the control block repeatedly presents two stimuli that differ only in one dimension, for example, tone. For instance, on each trial, listeners hear either /ta-high/ or /ta-mid/ then judge whether they heard the high tone or mid-tone. The orthogonal block contains an irrelevant dimensional variation, for example, vowel. On each trial, listeners hear /ta-high/, /ti-high/, /ta-mid/, or /ti-mid/ then judge whether they heard the high or mid tone. The AXB discrimination task adopts a similar

framework. On each trial in the control block, listeners hear three stimuli that differ only in one dimension (e.g., /ta-high/, /ta-high/, and /ta-mid/). They then pick the item that does not match the second item. On each trial in the orthogonal block, listeners hear three stimuli that differ in two dimensions (e.g., /to-high/, /ta-high/, and /ti-mid/). They then need to ignore the irrelevant variation (e.g., vowel) and make judgment based only on the relevant variation (e.g., tone). In both speeded classification and AXB tasks, the increase in response time in the orthogonal relative to the control block is known as Garner interference (Garner, 1974). Such interference has been largely taken to reflect integral processing of the relevant and irrelevant dimensions (for a review, see Algom & Fitoussi, 2016). By contrast, similar response time in both orthogonal and control blocks reflects independent processing of the two dimensions.

Tone language listeners perceive L1 tones and segmental information integrally, according to behavioral studies (e.g., Lin & Francis, 2014; Tong et al., 2008; Zou et al., 2017). Lin and Francis (2014) tested English and Mandarin listeners with speeded tone, consonant, and vowel classification tasks. Each speeded classification task consisted of a control (e.g., tone variations only) and two orthogonal blocks (e.g., tone + vowel variations; tone + consonant variations). Between-block comparisons showed that consonant variations increased the tone classification response time. This has reflected that Mandarin listeners cannot neglect irrelevant consonant variations while attending to tones (ibid.). This echoed a previous study that found interference effects between consonant, vowel, and tones in their respective speeded classifications (Tong et al., 2008). Notably, Tong et al. (2008) even reported an asymmetrical interference effect—vowels exerted stronger interference on tones than vice versa. However, this finding was not replicated in the later study (Lin & Francis, 2014). Subtle discrepancies aside, these studies support earlier claims that tone language listeners integrally process tones and segmental information (Lee & Nusbaum, 1993; Repp & Lin, 1990). Critically, behavioral tasks require overt responses from listeners. This has limited their ability to test perceptual integration at the preattentive level.

Mismatch negativity (MMN) additivity approach is a potent way to test perceptual integration at the preattentive level (Levänen et al., 1993; Paavilainen et al., 2003; Takegata et al., 1999). Typically, an MMN additivity experiment comprises two single deviant oddballs (e.g., tone deviant and vowel deviant) and a double deviant oddball (e.g., tone-and-vowel deviant). These oddballs elicit two single-MMNs (e.g., tone-MMN and vowel-MMN) and a double-MMN (e.g., tone-and-vowel-MMN). If the same neuronal population reacts to the two deviant features (e.g., tone and vowel) in the double deviant oddball, the neurons would underfire (Lidji et al., 2010). As such, the intensity of the neuronal firing would be lower than the sum of the neuronal firing in response to the single deviants, that is, MMN under additivity. Conversely, if the amplitude of the double-MMN is equal to the sum of the two single-MMNs, independent processing of the two deviant features is said to occur, that is, MMN additivity (Takegata et al., 1999). Source-localization further supports that MMN additivity reflects separate MMN generators for each deviant feature (Caclin et al., 2006; Takegata et al., 2001).

MMN additivity studies have reported preattentive tone-segmental integration in Cantonese (Choi et al., 2017; Yu et al., 2022) and Mandarin listeners (Gao et al., 2012). Choi et al. (2017) engaged Cantonese listeners in two single oddballs (tone deviant and vowel deviant) and a double oddball (tone-and-vowel deviant) with synthesized Cantonese syllables. The double-deviant MMN was smaller in amplitude than the sum of the single-deviant MMNs. This MMN nonadditivity reflected preattentive integral processing of Cantonese tones and vowels. Latency and topographic analyses

yielded consistent results. Despite the statistically robust results, Choi et al. (2017) only included a level tone contrast in the experiment, that is, high-level and mid-level tones. Recently, a similar MMN additivity study further extended the finding to a contour tone contrast, that is, high-rising and low-falling tones (Yu et al., 2022). Beyond tones and vowels, they also reported tone-consonantal integration.

Collectively, behavioral and neurophysiological evidence support tone-segmental integration in tone language listeners (Choi et al., 2017; Gao et al., 2012; Lin & Francis, 2014; Tong et al., 2008; Yu et al. 2022; Zou et al., 2017). On the one hand, these may reflect that the listeners are perceptually attuned to optimize processing for Chinese word form detection (Ho, 2015; Shen et al., 2021). On the other hand, one may argue that tone-segmental integration is universal—humans simply process f_0 and formant frequencies integrally regardless of language experience. Unlike in Chinese, f_0 has minimal lexical function in nontonal languages (Ladefoged & Johnson, 2011). Even in stress-timed Dutch and English, there are only few stress minimal pairs (Cutler, 2014). If English listeners are perceptually attuned to optimize processing for English word form detection, they should process f_0 and segmental information independently as separate information (Lee & Nusbaum, 1993).

Behavioral studies on non-tonal language listeners yielded conflicting results on the universality of tone-segmental integration (Lin & Francis, 2014; Zou et al., 2017; see also Lee & Nusbaum, 1993; Repp & Lin, 1990). In speeded tone and consonant classification tasks, English listeners showed similar response time across the baseline and orthogonal conditions (Lin & Francis, 2014). In other words, the English listeners could ignore irrelevant tonal/consonantal variations when judging consonants/tones. As such, the authors concluded that English listeners perceive tones and consonants independently. In a more recent year, Dutch listeners with varying Mandarin learning experience completed a modified AXB task (Zou et al., 2017). Consistent with Lin and Francis (2014), the authors claimed that Mandarin-naïve Dutch listeners perceived segmental and tonal information independently. However, nuanced analysis showed integral perception—the response time increased when their Dutch listeners had to ignore vowel variations when making tonal judgments. Consistent with earlier studies, this might imply the universality of tone-segmental integration—listeners perceive segmental and tonal information integrally regardless of L1 experience (Lee & Nusbaum, 1993; Repp & Lin, 1990). The present study can enrich the body of evidence in this theoretical issue by testing whether English listeners perceive segmental and tonal information integrally or independently.

Microscopic aspect: Segmental, prosodic, and lexical prosodic transfer

The microscopic aspect of cross-linguistic transfer concerns how L1 experience influences the perception and production of foreign phonemes (consonants and vowels), prosody (e.g., rhythm and intonation), and lexical prosody (e.g., tones and stress). Focusing on segmental transfer, the perceptual assimilation model theorizes how L1 phonological system determines the discriminability of nonnative phonemic contrasts (PAM; Best, 1995). The model posits that nonnative phonemes are categorized, uncategorized, or nonassimilated into L1 phonological categories, depending on their similarities with any L1 phonemes. The discriminability of a nonnative phonemic contrast is very high if the two phonemes are assimilated into separate L1 phonological categories (*two-category assimilation*) or if only one of the two phonemes is categorized (*uncategorized-categorized*). However, if two nonnative phonemes assimilate into the same L1 phonological category with equal goodness-of-fit, discriminability is low

(*single-category assimilation*). If two nonnative phonemes assimilate into the same L1 phonological category with unequal goodness-of-fit, discriminability varies from moderate to high depending on the goodness-of-fit difference (*category-goodness difference*). If two nonnative phonemes are uncategorized (*uncategorized-uncategorized*) or non-assimilated (*nonassimilable*), discriminability varies from low to high depending on their respective phonetic and acoustic similarities.

Regarding prosodic transfer, a prominent theoretical view is that L1 experience shapes L2 prosodic production (e.g., Gabriel & Kireva, 2014; Graham & Post, 2018; van Maastricht et al., 2016). In a previous study, Italian learners of Castilian Spanish produced Castilian Spanish that was rhythmically similar to Italian but not Castilian Spanish (Gabriel & Kireva, 2014). This suggested the transfer of L1 Italian rhythmic properties to L2 Spanish production. Linguistically, Dutch uses pitch accent distributions to mark focus while Spanish does not (e.g., Hualde, 2005; Rasier, 2006). In a Dutch speech production task, L1 Dutch speakers used pitch accent to mark focus but Spanish learners of Dutch did not (van Maastricht et al., 2016). Similarly, Dutch learners of Spanish produced pitch accent patterns in L2 Spanish that resembled Dutch instead of Spanish. Thus, besides speech rhythm, L1 shapes how L2 speakers prosodically mark information status (*ibid.*). Furthermore, the authors suggested that L2 proficiency modulated the transfer effects—less proficient L2 speakers tended to exhibit stronger prosodic transfer.

Concerning lexical prosody, most models and hypotheses focus on nonnative tone and stress perception. The perceptual assimilation model for suprasegmentals theorizes how L1 tonal or intonational system shapes nonnative tone perception (PAM-S; So & Best, 2010, 2014). Similar to PAM, PAM-S posits that nonnative listeners perceptually assimilate nonnative tones into their L1 tonal or intonational categories (c.f. Hallé et al., 2004). Like segmental discrimination, the discriminability of any given tone contrast hinges on the assimilation pattern (e.g., single-category, two-category, and uncategorized-categorized). Regarding nonnative stress perception, the cue-weighting transfer and acoustic-attentional-contextual hypotheses postulate that L1 experience shapes listeners' choice of acoustic cue for nonnative stress perception (Choi, 2022a; Chrabaszczyk et al., 2014; Kim & Tremblay, 2021). When multiple acoustic cues of stress are available (e.g., f_0 , duration, and intensity), L2 listeners would primarily utilize the cue with the largest lexical function in L1. For example, f_0 plays a lexical role in tone languages, so tone language listeners attend to it most heavily relative to duration and intensity for English stress discrimination (Choi et al., 2019; Wang, 2018). Gyeongsang-Korean listeners, L1 of whom has pitch accent features, also attend most heavily to f_0 for English stress sequence recall (Kim & Tremblay, 2021).

Macroscopic aspect: Dimensional transfer

Microscopically, the preceding models and hypotheses have theorized how L1 experience shapes the (a) discriminability of nonnative speech contrasts, (b) nonnative prosody production, and (c) the choice of acoustic cues for nonnative tone and stress perception (Best, 1995; Choi, 2022a; Kim & Tremblay, 2021; So & Best, 2010, 2014; van Maastricht et al., 2016). Aside from phonemes and acoustic cues, a holistic theory should address cross-linguistic transfer at the dimensional level. Multidimensional in nature, speech comprises source and filter (Ladefoged & Johnson, 2011; Tokuda, 2021). Tone language listeners perceive segmental and tonal information integrally in their L1 (Choi et al., 2017; Gao et al., 2012; Tong et al., 2008; Yu et al., 2022). Does their perceptual integrality transfer to a foreign language?

There is preliminary evidence supporting cross-linguistic transfer of perceptual integrality (Lin & Francis, 2014). As described previously, Mandarin listeners integrally processed segmental and tonal information in the Mandarin speeded classification tasks. In the same study, the authors added a context that drove the Mandarin listeners to process the Mandarin stimuli as English. After being told that the stimuli were English, the Mandarin listeners classified the stimuli into English consonants, vowels, and intonational categories. As in the Mandarin context, the Mandarin listeners showed Garner interference in the English context. Based on this finding, the authors have concluded that Mandarin listeners process segmental and suprasegmental information integrally even in English. This hints us that perceptual integrality can transfer to a foreign language.

Given methodological concerns, the preceding finding is only suggestive. In the previous study, the English stimuli were synthesized based on natural Mandarin productions by a Mandarin speaker (*ibid.*). Although their English stimuli may sound like English, they are Mandarin real words and produced as such by the speaker (e.g., /p^hi/ 批 *batch*; also sounds like *pea*). Even though the Mandarin listeners had been told that the stimuli were English words, one could not dismiss the possibility that the stimuli were processed as Mandarin words.

Overcoming this limitation, we use a set of unambiguously Thai stimuli produced by native Thai speakers. To avoid the Cantonese listeners from treating the Thai stimuli as Cantonese words, each Thai stimulus contained segmental information absent in Cantonese. For example, the consonant /t^h/ in the stimuli /t^hɔː-mid/, t^hɔː-low/, /t^hɔː-falling/, /t^hɔː-high/, and /t^hɔː-rising/ is absent in Cantonese. This also means that none of the stimuli is a real word in Cantonese, unlike in the Mandarin study (*ibid.*). In our study, Cantonese and English listeners discriminated Thai tones in the modified AX task. Conventionally, the AX task contains two stimuli on each trial (Choi & Chiu, 2022; McGuire, 2010). Listeners then judge as quickly as possible whether the two stimuli are different. Our modified AX task contains a control and an orthogonal block. In the control block, the segmental information was constant within each trial (e.g., /t^hɔː-mid/-/t^hɔː-mid/ for AA; /t^hɔː-mid/-/t^hɔː-rising/ for AB). In the orthogonal block, the segmental information varied (e.g., /t^hɔː-mid/-/k^hu-mid/ for AA; /t^hɔː-mid/-/k^hu-rising/ for AB). If perceptual integrality transfers from Cantonese to Thai, the Cantonese listeners should show an increase in response time and/or decrease in accuracy in the orthogonal block relative to the control block.

Research questions

In short, the present study aims to address these research questions:

1. Do Cantonese listeners integrally or independently perceive Thai tones and segmental information?
2. Do English listeners integrally or independently perceive Thai tones and segmental information?

Methods

Participants

To abide by the social distancing rules associated with COVID-19, the experiment switched from face-to-face to online. We recruited 30 Cantonese and 30 English listeners through e-mail and an online platform (Prolific; www.prolific.co). According

to power analysis, a sample size of 60 would achieve the power of .95 given a medium effect size and $r = .46$ among the repeated measures obtained in the pilot study. After signing up for the study, the participants completed an initial language background screening on phone, e-mail, or Prolific. All English listeners (a) were living in an English-speaking country, (b) learned and spoke English since birth as a first language, (c) had not learned any tone language, and (d) had no reported hearing difficulties. All Cantonese listeners (e) were living in Hong Kong, (f) learned and spoke Cantonese as a first language, (g) had not learned Thai, and (h) had no reported hearing difficulties.

The experiment was conducted on an online platform (Gorilla; Anwyl-Irvine et al., 2020; Choi, 2021a, 2022b; Jasmin et al., 2021). An automatic procedure was applied to block any access through tablets and smart phones. As such, all participants completed the experiments with a desktop or laptop computer. The participants were asked to sit comfortably in a quiet environment and wear headphones/earphones. To monitor attention, five attention-check trials were pseudorandomly embedded throughout the experiment. Each attention-check trial contained two identical audio stimuli (see Supplementary Materials). On each attention-check trial, participants essentially judged whether the two identical stimuli sounded the same. Because the two stimuli were acoustically identical, the attention-check trials could be answered easily. To be empirically stringent, only one mistake was allowed, that is, minimum 80% accuracy. As such, four Cantonese listeners were excluded from the dataset. Moreover, one Cantonese listener was rejected by the system for exceeding the 95-minute experimental time limit. The final sample had 25 Cantonese (8 males, 17 females; mean age = 25.04 years; $SD = 5.66$ years) and 30 English listeners (15 males, 14 females, 1 unknown; mean age = 27.77 years; $SD = 6.30$ years). The Cantonese and English groups had very high accuracies on the attention check trials ($M_{Cantonese} = 97\%$, $SD_{Cantonese} = 7\%$; $M_{English} = 96\%$, $SD_{English} = 8\%$).

Stimuli

Two native Thai speakers (1 male and 1 female) participated in the stimuli recording. They naturally produced five Thai tones (mid, low, falling, high, and rising) embedded in four syllables (/te^hɔː/, /k^huː/, /t^hiːa/, and /wuːa/). As described previously, these syllables are phonotactically legal in Thai but not in Cantonese. Altogether, there were 40 stimuli (5 tones × 4 syllables × 2 genders). Figures 1, S1, S2, and S3 summarize the f0 profiles of the stimuli.

Stimuli presentation

The modified AX task contained a control block and an orthogonal block. On each trial, two audio stimuli were presented with an interstimulus interval of 500 ms. To prevent listeners from making simple acoustic comparisons, A and X were always produced by speakers of different genders. The gender order of A and X was counterbalanced across trials. In the control block, the segmental information was constant within each trial, for example, /te^hɔː-mid/—/te^hɔː-mid/ for AA and /te^hɔː-mid/—/te^hɔː-low/ for AB. Listeners indicated on a keyboard whether the two stimuli contained the same [f] or different [j] tones. Accuracy and response time were recorded on each trial. The control block contained all the 10 Thai tone contrasts (mid-low, mid-falling, mid-high, mid-rising, low-falling, low-high, low-rising, falling-high, falling-rising, and high-rising tones) embedded in four syllables (/te^hɔː/, /k^huː/, /t^hiːa/, and /wuːa/), yielding 40 AB trials.

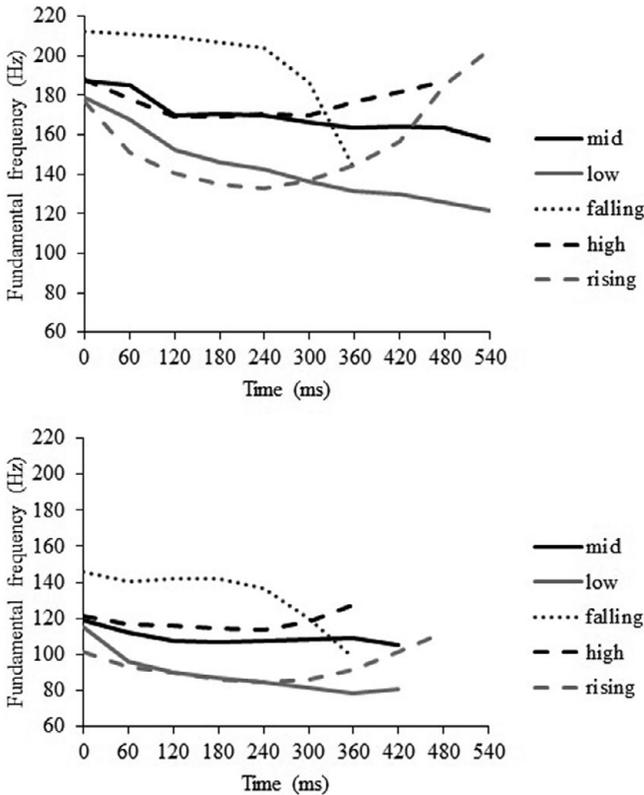


Figure 1. The F0 profiles of the female- (top) and male- (bottom) produced Thai tones in the syllable /tʰi:a/. Note: The first detectable f0 value of each stimulus is at the zero-time point.

To balance the trial types, there were 40 AA trials (mid-mid, low-low, falling-falling, high-high, and rising-rising tones) embedded in the same four syllables. In total, the control block had 80 experimental trials. Participants first completed three practice trials with feedback. The internal consistency in the control block was high (Cronbach's $\alpha = .85$).

The orthogonal block had the same procedure as the control block. The only difference was that the segmental information varied within each trial, for example, /te^hɔ:-mid/—/k^hu:-mid/ for AA and /k^hu:-mid/—/te^hɔ:-low/ for AB. As in the control block, the number of occurrence of each syllable (/te^hɔ:/, /k^hu:/, /tʰi:a/, and /wu:a/) was equal in the orthogonal block. The order of presentation of the stimuli was also counterbalanced. Listeners were told to ignore the segmental variations when making the tone judgments. Like the control block, the orthogonal block contained the 10 Thai tone contrasts that were spread across 80 experimental trials. The block began after three practice trials with feedback. The internal consistency in the orthogonal block was high (Cronbach's $\alpha = .80$).

Supplementary face-to-face experiment

To address the potential concern about the quality of the online data, we retrospectively tested 13 Cantonese listeners (6 males, 7 females; mean age = 26.62 years, $SD = 6.80$ years)

and 1 English listener (female; age = 22 years) who participated in another study. We could only recruit one English listener due to the cancellation of foreign exchange programs in response to the COVID-19 pandemic. The face-to-face experiment took place in a sound booth at the University of Hong Kong. All listeners heard the stimuli using Sennheiser HD280 PRO headphones in the same physical environment. The Cantonese and English listeners met their respective participant inclusion criteria as the online experiment.

All face-to-face participants completed the modified AX task with the same procedure as previously mentioned. To be even more empirically stringent, we doubled the number of trials of the control and orthogonal blocks. This yielded 160 experimental trials per block (80 AB and 80 AA trials). In each block, we randomly inserted three attention-check trials, all of which were answered correctly ($M = 100\%$, $SD = 0\%$). As in the online experiment, we counterbalanced the order of block across participants. Like the online data, the internal consistency was high in the control (Cronbach's $\alpha = .91$) and orthogonal blocks (Cronbach's $\alpha = .93$).

Results

Main analysis

Linear mixed-effects analyses were performed in the R statistical language (Version 4.1.3; R Core Team, 2020) using the lme4 package (Bates et al., 2015). The lmerTest package (Kuznetsova et al., 2017) was used to calculate the p -values of the models. Our predictor variables were block and language group. Block (control vs. orthogonal) and language group (Cantonese vs. English) were categorical with two levels, with Cantonese control as the reference level. Post-hoc analyses were conducted through pairwise comparisons using the emmeans package (Length, 2022).

Accuracy

We ran separate models to explore two measures of accuracy: (a) discrimination sensitivity index d' and (b) proportion of accuracy. In each block of the modified AX task, the d' was computed by subtracting the z -transform of the hit rate by that of the false alarm rate (signal detection theory; MacMillan & Creelman, 2005). Each correct AB trial was regarded as a hit whereas each incorrect AA trial was regarded as a false alarm. As in previous studies, we manually adjusted perfect false alarm (0) and hit rates (1.0) to .01 and .99, respectively, to avoid infinity d' (Bidelman et al., 2013; Choi et al., 2019). On average, the Cantonese listeners showed a d' of 1.90 in the control block ($SD = .61$; range = .91–3.25) and a d' of .91 in the orthogonal block ($SD = .56$; range = .13–2.28); whereas the English listeners showed a d' of .71 in the control block ($SD = .54$; range = -.34–1.78) and a d' of .76 in the orthogonal block ($SD = .74$; range = -.79–2.08).

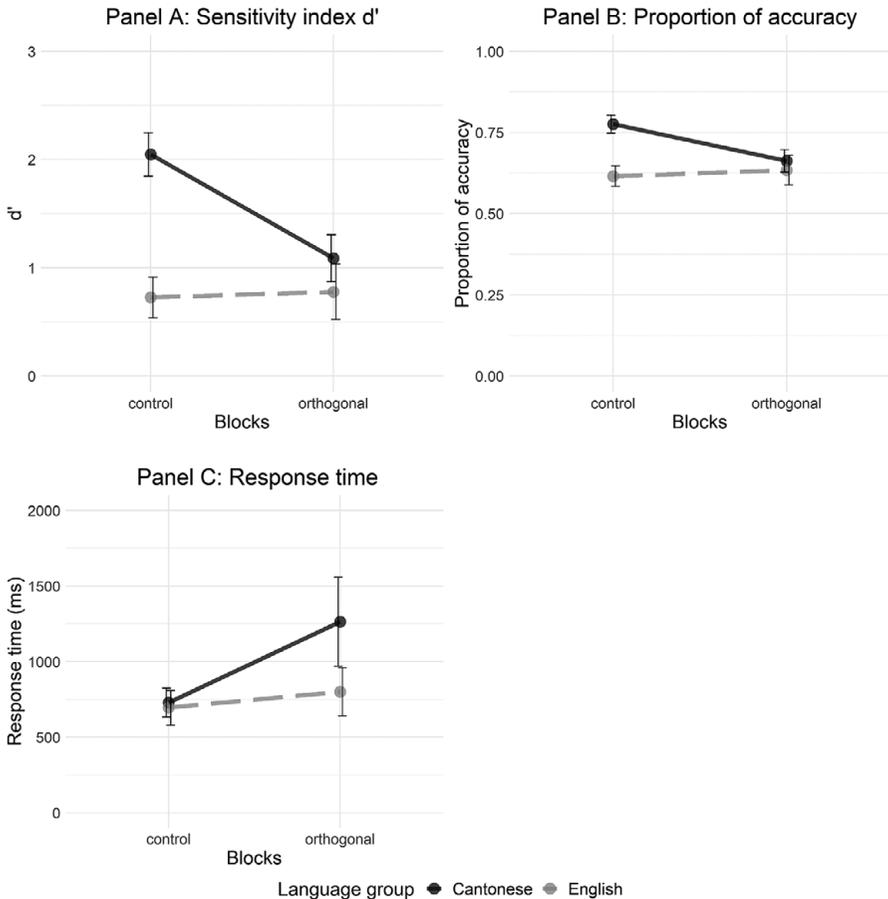
A linear mixed-effects analysis was run on d' , with block and language group as fixed effects and participant as a random intercept. The model specification was as follows:

$$d \sim \text{block} * \text{language_group} + (1|\text{participant})$$

Table 1 shows the coefficient estimates from this model and Figure 2 Panel A visualizes this model. There was a significant main effect of block with a negative coefficient, suggesting that the Cantonese listeners' d' was lower in the orthogonal block than in the control block. There was also a significant main effect of language group, suggesting

Table 1. Coefficient estimates from the mixed-effects models predicting discrimination accuracy

	Sensitivity index d'				Proportion of accuracy			
	Estimate	SE	t	p	Estimate	SE	z	p
intercept	1.896	.124	15.297	< .001	1.196	.257	4.655	< .001
block	-.985	.136	-7.249	< .001	-.758	.241	-3.147	< .01
language group	-1.189	.168	-7.084	< .001	-.944	.129	-7.298	< .001
block * language group	1.041	.184	5.662	< .001	.811	.155	5.224	< .001

**Figure 2.** Averages by block and language group, with Panel A illustrating d' , Panel B illustrating the proportion of accuracy, and Panel C illustrating the response time.

Note: The solid black line represents the Cantonese listeners, and the dashed gray line represents the English listeners. Error bars indicate 95% confidence intervals.

that the Cantonese listeners were more accurate than the English listeners in the control block. Moreover, the interaction between block and language group was significant. Pairwise comparisons with Bonferroni corrections revealed that the Cantonese listeners had a significantly higher d' in the control block than in the orthogonal block

(estimate = .985, $SE = .136$, $t = 7.249$, $p < .001$), whereas the English listeners had similar d 's across the control and orthogonal blocks (estimate = $-.057$, $SE = .124$, $t = -.457$, $p = .650$). In other words, segmental variations impeded Thai tone discrimination among the Cantonese but not the English listeners.

The second accuracy measure was the proportion of trials in each block to which listeners gave a correct response. For each correct trial, a point of 1 was given; otherwise, a point of 0 was given. On average, the Cantonese listeners showed a mean accuracy of .78 in the control block ($SD = .07$; range = .59–.91) and .66 in the orthogonal block ($SD = .09$; range = .53–.86). Meanwhile, the English listeners showed a mean accuracy of .62 in the control block ($SD = .09$; range = .44–.81) and .63 in the orthogonal block ($SD = .13$; range = .35–.85).

A logistic mixed-effects analysis was run with proportion of accuracy as the dependent variable. Block and language group were again entered as fixed effects. A random slope of block by participants and random intercepts of item, tone pair, and syllable pair were entered. The model specification was as follows:

$$\text{proportion_trial} \sim \text{block} * \text{language_group} + (1 + \text{block} | \text{participant}) + (1 | \text{item}) \\ + (1 | \text{tone_pair}) + (1 | \text{syllable_pair})$$

Table 1 summarizes the coefficient estimates from this model and Figure 2 Panel B visualizes this model. As in the d' analysis, we observed significant main effects of block and language group, as well as a significant interaction. Overall, pairwise comparisons suggested that the Cantonese listeners showed a significantly higher proportion of accuracy in the control block than in the orthogonal block (estimate = .758, $SE = .241$, $z = 3.147$, $p < .01$). However, the English listeners showed similar proportion of accuracy across the two blocks (estimate = $-.053$, $SE = .234$, $z = -.228$, $p = .820$). This pattern once again showed that segmental variations hindered the Cantonese but not the English listeners' Thai tone discrimination.

Response time

On average, the Cantonese listeners demonstrated response time of 729 ms in the control block ($SD = 244.29$; range = 271–1287) and 1264 ms in the orthogonal block ($SD = 748.63$; range = 320–3087). However, the English listeners showed a response time of 695 ms in the control block ($SD = 323.13$; range = 365–1780) and 799 ms in the orthogonal block ($SD = 445.61$; range = 340–2141).

Response time was log-transformed to correct for skewness. The log-transformed response time was analyzed with a linear mixed-effects model, with block and language group entered as fixed effects. A random slope of block by participants and random intercepts of item, tone pair, and syllable pair were entered. The model specification was as follows:

$$\log(\text{rt}) \sim \text{block} * \text{language_group} + (1 + \text{block} | \text{participant}) + (1 | \text{item}) \\ + (1 | \text{tone_pair}) + (1 | \text{syllable_pair})$$

Table 2 shows the coefficient estimates from this model, and Figure 2 Panel C visualizes this model. The effect of block with a significant positive coefficient indicated that the Cantonese listeners showed a longer response time in the orthogonal block than in the control block. While the effect of language group was not significant, the interaction

Table 2. Coefficient estimates from the linear mixed-effects model predicting response time

	Estimate	SE	<i>t</i>	<i>p</i>
intercept	6.205	.104	59.527	< .001
block	.398	.104	3.840	< .001
language group	.025	.103	.248	.805
block * language group	-.347	.100	-3.478	< .01

between block and language group suggested a possible attenuation of this pattern for the English listeners compared to the Cantonese listeners. Pairwise comparison revealed that the Cantonese listeners took longer to respond to the stimuli in the orthogonal block than in the control block (estimate = $-.398$, $SE = .104$, $t = -3.840$, $p < .001$). However, the English listeners showed similar response time across the orthogonal and control blocks (estimate = $-.051$, $SE = .099$, $t = -.517$, $p = .609$). Consistent with the accuracy analyses (d' and proportion of accuracy), segmental variations worsened the performance of the Cantonese but not the English listeners.

Supplementary analysis of face-to-face data

We analyzed the face-to-face data to supplement the previously mentioned online results. Table 3 and Figure 3 show the descriptive statistics of the accuracy measures and the response time of the face-to-face group. Because there was not enough face-to-face data from English listeners, we limited the following statistical models to data from the Cantonese listeners. Specifically, we asked whether the face-to-face Cantonese listeners integrally process Thai tones and segmental information.

Accuracy

Following the strategies used in the main analysis, we ran separate models to explore d' and proportion of accuracy as measures of accuracy.

A pairwise t -test on d' revealed a significant block difference, where the face-to-face Cantonese listeners showed lower d' in the orthogonal block than in the control block ($t(12) = 7.108$, $p < .001$).

A logistic mixed-effects analysis was run on proportion of accuracy, with block as a fixed effect. A random slope of block by participants and random intercepts of item, tone pair, and syllable pair were entered. The model specification was as follows:

$$\text{proportion_trial} \sim \text{block} + (1 + \text{block} | \text{participant}) + (1 | \text{item}) + (1 | \text{tone_pair}) \\ + (1 | \text{syllable_pair})$$

Table 3. Descriptive statistics of the accuracy measures (d' score and proportion of accuracy) and response time from the face-to-face data

	Cantonese listeners (N = 13)		English listener (N = 1)	
	Control	Orthogonal	Control	Orthogonal
D' score	2.34 ($SD = .59$)	1.43 ($SD = .78$)	1.26	1.20
Proportion of accuracy	.85 ($SD = .08$)	.74 ($SD = .12$)	.73	.73
Response time (ms)	678 ($SD = 272$)	1278 ($SD = 798$)	990	1036

Note: No SD s are reported for the English group as we only have data from one English listener.

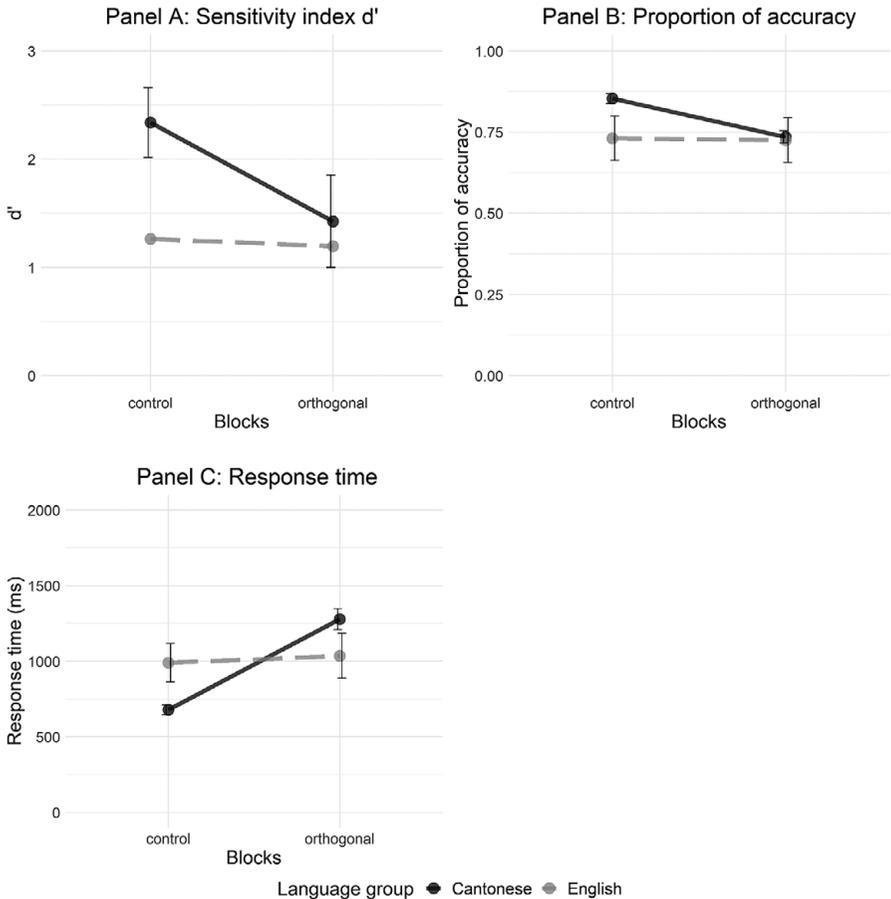


Figure 3. Discrimination performance of the face-to-face participants by language group, with Panel A illustrating d' , Panel B illustrating the proportion of accuracy, and Panel C illustrating response time. *Note:* The solid black line represents the Cantonese listeners, and the dashed gray line represents the English listeners. Error bars indicate 95% confidence intervals. There are no error bars for the English group's d' score as we only have data from one English listener.

Consistent with the d' analysis, there was a significant main effect of block where the Cantonese listeners' accuracy was lower in the orthogonal block than in the control block (estimate = -1.031 , $SE = .257$, $z = -4.015$, $p < .001$). Like the online Cantonese listeners, the face-to-face Cantonese listeners discriminated Thai tones less accurately in the orthogonal than in the control block.

Response time

The log-transformed response time was analyzed with a linear mixed-effects model, with block entered as fixed effects. A random slope of block by participants and random intercepts of item, tone pair, and syllable pair were entered. The model specification was as follows:

$\log(\text{rt}) \sim \text{block} + (1 + \text{block}|\text{participant}) + (1|\text{item}) + (1|\text{tone_pair}) + (1|\text{syllable_pair})$

There was a significant effect of block, suggesting that the Cantonese listeners responded faster in the control block than in the orthogonal block (estimate = .505, $SE = .092$, $t = 5.464$, $p < .001$). Therefore, segmental variations worsened the face-to-face Cantonese listeners' discrimination performance.

Summary of findings

Our main analysis showed that segmental variations impeded Thai tone discrimination among the Cantonese but not the English listeners. While the English listeners showed consistent performance across the blocks, the Cantonese listeners were less accurate and took longer to respond in the orthogonal block than in the control block. Analysis of the face-to-face data from Cantonese listeners yielded the same pattern. Overall, segmental variations made Thai tone discrimination more difficult for the Cantonese listeners.

Discussion

The present study examines whether (a) Cantonese and (b) English listeners integrally or independently perceive Thai tones and segmental information. The results are simple and straightforward. Relative to their own performance in the control block, the Cantonese listeners showed lower accuracy and longer response time in the orthogonal block. The observed interference effect of segmental variation reflects tone-segmental integrality, consistent with the previous finding from Mandarin listeners (Lin & Francis, 2014). Using unambiguously foreign (Thai) stimuli, the present study has overcome the previous methodological limitation and offers stronger support for the claim—tone language listeners show tone-segmental integrality in foreign speech perception.

Is tone-segmental integrality language specific (i.e., limited to tone language listeners) or universal (i.e., applicable to all humans)? In our modified AX discrimination task, the English listeners showed similar accuracies and response time across the control and the orthogonal blocks. Unlike the early studies, the present results suggest that English listeners perceive tone and segmental information independently (Lee & Nusbaum, 1993; Repp & Lin, 1990). This finding supports the language specific view but not the universal view. For the Cantonese listeners, this further indicates that their tone-segmental integrality in Thai is transferred from L1 (rather than being an innate perceptual mode of humans).

Dimensional transfer hypothesis

The preceding findings motivate the dimensional transfer hypothesis, which conceptualizes cross-linguistic transfer at the macroscopic (dimensional) level. When multiple acoustic cues are available, the cue-weighting transfer hypothesis posits that L2 listeners attend most heavily the one with the largest L1 functional importance (see Kim & Tremblay, 2021). Taking a similar perspective, our dimensional transfer hypothesis postulates that L1 functional importance shapes segmental-suprasegmental integrality (or nonintegrality) in foreign speech perception. Phonologically, segmental and tonal information are essential parts of Cantonese lexical representations (Shen et al., 2021).

To optimize the perceptual system for L1 word recognition, Cantonese listeners perceive both dimensions integrally as the same kind of information (Choi et al., 2017; Yu et al., 2022). The hypothesis posits that such perceptual mode is transferrable to foreign speech. Even though the Thai tones and segmental information do not exist in Cantonese listeners' phonological repertoire, the Cantonese listeners perceived them integrally. As such, the hypothesis further postulates that such transfer is not phonetically driven but rather experience driven. Unlike the cue-weighting transfer hypothesis and other microscopic theoretical accounts, what is being transferred across language is not any specific L1 phoneme/tone/acoustic cue but the cognitive experience of integrating segmental and tonal dimensions (e.g., Best, 1995; Kim & Tremblay, 2021; So & Best, 2010, 2014).

The dimensional transfer hypothesis also applies to foreign speech perception in nontonal languages. Despite the aforementioned limitation of the stimuli, Lin and Francis (2014) has provided initial evidence that Mandarin listeners perceive English segmental and f_0 information integrally. In a recent active oddball task, Cantonese and English listeners detected English stress deviants in (a) a segmentally varying context and (b) a segmentally nonvarying context (Choi, 2021b). Though no such effect was found in the English listeners, segmental variation decreased the behavioral accuracy of the Cantonese listeners. This has indicated that Cantonese listeners perceive English stress and segmental information integrally. As mentioned previously, suprasegmental information has small lexical role in English than in Cantonese (Cutler, 2014; Shen et al., 2021). However, because segmental-suprasegmental integration is lexically functional in Cantonese, Cantonese listeners perceive both dimensions as the same kind of information even in English. The cross-linguistic transfer of segmental-suprasegmental integrality fits the notion of the dimensional transfer hypothesis.

On top of tone language listeners, the dimensional transfer hypothesis applies to nontonal language listeners. In English, suprasegmental information has minimal lexical function (Cutler, 2014). To optimize processing for English word recognition, English listeners process segmental and suprasegmental information independently as separate information (Lin & Francis, 2014; cf. Repp & Lin, 1990). In the present study, the English listeners transferred the perceptual nonintegrality from L1 English to a foreign tone language, that is, Thai. Simply put, nonintegrality is the optimal processing mode in English, so the English listeners adopt the same for Thai language. This reinforces the hypothesis that L1 functional importance shapes the integrality or nonintegrality of foreign segmental and suprasegmental speech perception.

The dimensional transfer hypothesis is open to the idea that L2 tone language learning can help establish perceptual integrality. In a modified AXB task, advanced Dutch learners of Mandarin integrally perceived Mandarin tone and segmental information, unlike Mandarin-naïve Dutch listeners who showed independent perception (Zou et al., 2017). This seems to indicate that L2 learning experience can alter the perceptual mode. Crucially, beginning Dutch learners of Mandarin showed independent perception of Mandarin tone and segmental information, like those without Mandarin experience (ibid.). This has suggested that the alteration of perceptual mode requires *long-term* L2 learning experience. Contrasting the Dutch findings, advanced Cantonese learners of English integrally perceive English stress and segmental information, unlike native English listeners who showed independent perception (Choi, 2021b). To our knowledge, there is as yet no report that tone language listeners show segmental-suprasegmental nonintegrality in nontonal language. This may imply that perceptual nonintegrality is more malleable than perceptual integrality. This possible interpretation awaits further empirical scrutiny.

Methodological implications

The present study has designed a novel paradigm for examining perceptual integrality—the modified AX task. To date, most studies on perceptual integrality (e.g., Lin & Francis, 2014; Repp & Lin, 1990; Tong et al., 2008) adopted Garner's (1974) speeded classification task. This limited most integrality findings to speech classification. More recently, Zou et al. (2017) designed a modified AXB task that yielded interference effects between segmental and tonal information. This extended previous findings to speech discrimination. In the present study, the modified AX task also yielded the interference effect of segmental variation on tone discrimination. Unlike in the previous studies, this interference effect manifested not only in response time but also in proportion of accuracy and sensitivity index (Lin & Francis, 2014; Repp & Lin, 1990; Tong et al., 2008; Zou et al., 2017). In other words, the modified AX task has demonstrated that segmental variation not only slows down perceptual processing but also decreases perceptual sensitivity. As such, the modified AX task appears to be more sensitive than speeded classification and modified AXB tasks in assessing perceptual integrality.

Conclusion

In conclusion, the present results reflect that Cantonese listeners integrally perceive Thai tones and segmental information whereas English listeners independently perceive them. Based on past and present findings, we propose the dimensional transfer hypothesis (Choi, 2021b; Lin & Francis, 2014; Zou et al., 2017). This hypothesis posits that perceptual integrality is transferrable across languages. Specifically, L1 functional importance shapes the perceptual integrality (or nonintegrality) of foreign suprasegmental and segmental information. The hypothesis further assumes that the transfer is driven by the cognitive experience in processing segmental and suprasegmental dimensions integrally/independently in L1, rather than individual phonemes or tones, per se. Futures studies can elucidate the malleability of perceptual integrality and nonintegrality in response to different amounts of L2 learning experience. Methodologically, the present study has developed the modified AX task, which is a potent way to examine perceptual integrality.

Supplementary Materials. To view supplementary material for this article, please visit <http://doi.org/10.1017/S0272263122000511>.

Acknowledgments. This work was supported by the University of Hong Kong (Seed Fund for Basic Research for New Staff [202107185043] and Project-Based Research Funding [supported by Start-up Fund]). We appreciate Shirlyn Chiu, Sze Hung Chung, Kin Chak Lau, and Mei Sze Wong for their assistance in data collection and entry.

Competing Interests. The authors declare none.

Data Availability Statement The experiment in this article earned Open Data and Open Materials badges for transparent practices. The materials and data are available at https://osf.io/82697/?view_only=245529a3b90540f89629f36d9b79bec1

References

- Algom, D., & Fitousi, D. (2016). Half a century of research on Garner interference and the separability–integrality distinction. *Psychological Bulletin*, 142, 1352–1383.
- Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52, 388–407.

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Bayne, T., & Chalmers, D. J. (2003). *What is the unity of consciousness?* (A. Cleeremans, Ed.). Oxford University Press.
- Best, C. T. (1995). *A direct realist view of cross-language speech perception* (W. Strange, Ed.). York Press.
- Bidelman, G. M., Hutka, S., & 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.
- Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M.-H., & McAdams, S. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, 18, 1959–1972.
- Chang, C. B. (2018). Perceptual attention as the locus of transfer to nonnative speech perception. *Journal of Phonetics*, 68, 85–102.
- Chiba, T., & Kajiyama, M. (1942). *The vowel: Its nature and structure*. Tokyo-Kaiseikan.
- Choi, W. (2021a). Musicianship influences language effect on musical pitch perception. *Frontier in Psychology*, 12, 712–753.
- Choi, W. (2021b). Cantonese advantage on English stress perception: Constraints and neural underpinnings. *Neuropsychologia*, 158, 107888.
- Choi, W., & Chiu, M. M. (2022). Why aren't all Cantonese tones equally confusing to English listeners? *Language and Speech*. Advance online publication. doi:10.1177/00238309221139789
- Choi, W. (2022a). Theorizing positive transfer in cross-linguistic speech perception: The acoustic-attentional-contextual hypothesis. *Journal of Phonetics*, 91, 101135.
- Choi, W. (2022b). What is “music” in music-to-language transfer? Musical ability but not musicianship supports Cantonese listeners' English stress perception. *Journal of Speech, Language, and Hearing Research*, 65, 4047–4059. doi:10.1044/2022_JSLHR-22-00175
- Choi, W., Tong, X., & Samuel, A. G. (2019). Better than native: Tone language experience enhances second language English lexical stress discrimination in Cantonese-English bilinguals. *Cognition*, 189, 188–192.
- Choi, W., Tong, X., Gu, F., Tong, X., & Wong, L. (2017). On the early neural perceptual integrality of tones and vowels. *Journal of Neurolinguistics*, 41, 11–23.
- Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech Language and Hearing Research*, 57, 1468–1479.
- Cutler, A. (2014). *Native listening, language experience and the recognition of spoken words*. The MIT Press.
- Gabriel, C., & Kireva, E. (2014). Speech rhythm and intonation of Buenos Aires Spanish and L2 Castilian Spanish produced by Italian native speakers. *Studies in Second Language Acquisition*, 36, 257–281.
- Gao, S., Hu, J., Gong, Q., Chen, S., Kendrick, K. M., & Yao, D. (2012). Integration of consonant and pitch processing as by the absence of additivity in mismatch negativity. *PLoS One*, 7, e38289.
- Garner, W. R. (1974). *The processing of information and structure*. Lawrence Erlbaum Associates.
- Graham, C., & Post, B. (2018). Second language acquisition of intonation: Peak alignment in American English. *Journal of Phonetics*, 66, 1–14.
- Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32, 395–421.
- Ho, A. (2015). Electrophysiological evidence for the integral nature of tone in Mandarin spoken word recognition [Master dissertation, McMaster University]. McMaster University Research Repository. https://macsphere.mcmaster.ca/bitstream/11375/18281/2/Ho_Amanda_SW_2015September_MSc.pdf
- Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge University Press.
- Jasmin, K., Sun, H., & Tierney, A. T. (2021). Effects of language experience on domain. *Journal of Experimental Psychology*, 149, 914–934.
- Kim, H., & Tremblay, A. (2021). Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. *Journal of Phonetics*, 87, 101059.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26.
- Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (6th ed.). Wadsworth.
- Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics*, 53, 157–165.

- Length, R. V. (2022). emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.4-1. <https://CRAN.R-project.org/package=emmeans>
- Levänen, S., Hari, R., McEvoy, L., & Sams, M. (1993). Responses of the human auditory cortex to changes in one versus two stimulus features. *Experimental Brain Research*, 97, 177–183.
- Lidji, P., Jolicour, P., Kolinsky, R., Moreau, P., Connolly, J. F., & Peretz, I. (2010). Early integration of vowel and pitch processing: A mismatch negativity study. *Clinical Neurophysiology*, 121, 533–541.
- Lin, M., & Francis, A. L. (2014). Effects of language experience and expectations on attention to consonants and tones in English and Mandarin Chinese. *Journal of the Acoustical Society of America*, 136, 2827–2838.
- MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New York: Psychological Press.
- McGuire, G. (2010). *A brief primer on experimental designs for speech perception research*. (Tech. Rep.). Department of Linguistics, UC Santa Cruz.
- Paavilainen, P., Mikkonen, M., Kilpeläinen, M., Lehtinen, R., Saarela, M., & Tapola, L. (2003). Evidence for the different additivity of the temporal and frontal generators of mismatch negativity: A human auditory event-related potential study. *Neuroscience Letter*, 349, 79–82.
- Rasier, L. (2006). *Prosodie en vreemdetaalverwerving: accentdistributie in het Frans en in het Nederlands als vreemde taal* [Prosody and second language acquisition: Accent distribution in French and Dutch as a foreign language]. Unpublished doctoral dissertation, Université catholique de Louvain, Belgium
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Repp, B. H., & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception: a cross-linguistic study. *Journal of Phonetics*, 18, 481–495.
- Revonsuo, A., & Newman, J. (1999). Binding and consciousness. *Consciousness and Cognition*, 8, 123–127.
- Shen, W., Hyönä, J., Wang, Y., Hou, M., & Zhao, J. (2021). The role of tonal information during spoken-word recognition in Chinese: Evidence from a printed-word eye-tracking study. *Memory and Cognition*, 49, 181–192.
- So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, 53, 273–293.
- 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.
- Takegata, R., Huottilainen, M., Rinne, T., Näätänen, R., & Winkler, I. (2001). Changes in acoustic features and their conjunctions are processed by separate neuronal populations. *Neuroreport*, 12, 525–529.
- Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letter*, 266, 109–112.
- Tokuda, I. T. (2021). *The source/filter theory of speech*. Oxford University Press.
- Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognition*, 23, 689–708.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- van Maastricht, L., Krahmer, E., & Swerts, M. (2016). Prominence patterns in a second language: Intonational transfer from Dutch to Spanish and vice versa. *Language Learning*, 66, 124–158.
- Wang, Y. (2018). *Native Mandarin speakers' production and perception of English voiceless interdental and alveolar fricatives*. LSHK Annual Research Forum 2018, City University of Hong Kong, Hong Kong.
- Yip, M. (2002). *Tone*. Cambridge University Press.
- Yu, K., Chen, Y., Wang, M., Wang, R., & Li, L. (2022). Distinct but integrated processing of lexical tones, vowels, and consonants in tonal language speech perception: Evidence from mismatch negativity. *Journal of Neurolinguistics*, 61, 101039.
- Zou, T., Chen, Y., & Caspers, J. (2017). The developmental trajectories of attention distribution and segment-tone integration in Dutch learners of Mandarin tones. *Language and Cognition*, 20, 1017–1029.

Cite this article: Choi, W. and Tsui, R. K. -Y. (2023). Perceptual integrality of foreign segmental and tonal information: Dimensional transfer hypothesis. *Studies in Second Language Acquisition*, 45: 1056–1073. <https://doi.org/10.1017/S0272263122000511>