Cross-Domain Effects of Tone Categorical Perception Patterns and Tone Complexity on Musical Pitch Perception

Authors: Li Xianzhuo, Xiao Rong, Liang Dandan, Liang Dandan

Date: 2022-03-06T12:13:47+00:00

Abstract

Using pitch identification and discrimination tasks, with native speakers of Chinese, Vietnamese, and Russian as participants, this study investigated two language processing-level factors, namely tonal categorical perception patterns and differences in complexity across tonal systems of different languages, and their influence on cross-domain musical pitch perception. The results showed: (1) The auditory discrimination results of the two tonal language groups (Chinese and Vietnamese) conformed to a categorical perception pattern, whereas native Russian speakers exhibited continuous perception. Under both linguistic and musical stimulus conditions, no significant differences were observed in tonal language native speakers across measures including category boundary width, within-category discrimination rate, between-category discrimination rate, and discrimination peak. (2) No significant difference was found in the musical pitch discrimination task results between the two tonal language groups (Chinese and Vietnamese). The experimental results indicate that, at the behavioral level, native tonal categorical perception patterns can transfer to musical pitch perception, but a complex tonal system does not facilitate fine-grained cross-domain musical pitch perception. The experimental results support the "shared processing" hypothesis from the perspective of language' s influence on musical pitch processing.

Full Text

The Cross-Domain Influence of Tonal Categorical Perception and Tonal Complexity on Musical Pitch Perception

LI Xianzhuo[1,2], XIAO Rong[2], LIANG Dandan[2]
[1] International College for Chinese Studies, Nanjing Normal University, Nanjing 210097, China

[2] School of Chinese Language and Literature, Nanjing Normal University, Nanjing 210097, China

Abstract

This study employed pitch identification and discrimination tasks to investigate how two language-processing factors—tonal categorical perception patterns and differences in tonal system complexity across languages—affect cross-domain musical pitch perception. Native speakers of Mandarin Chinese, Vietnamese, and Russian served as participants. The results revealed: (1) Both Chinese and Vietnamese tonal language speakers exhibited categorical perception patterns in their listening discrimination results, whereas Russian speakers showed continuous perception. No significant differences were observed between language and music stimulus conditions for tonal language speakers across metrics including category boundary width, within-category discrimination rate, between-category discrimination rate, and discrimination peak. (2) No significant differences emerged between the Chinese and Vietnamese tonal language groups in musical pitch discrimination task performance. These findings demonstrate that at the behavioral level, native tonal categorical perception patterns can transfer to musical pitch perception, yet complex tonal systems do not facilitate fine-grained cross-domain musical pitch perception. The results support the "shared domain-general view" from the perspective of language' s influence on musical pitch processing.

Keywords: categorical perception, pitch, tone, music
Classification Number: B842

Pitch constitutes a fundamental acoustic attribute shared by language and music. In language, pitch manifests as prosodic information encompassing intensity, intonation, and lexical tone. Tone represents a unique linguistic phenomenon wherein tonal languages employ different pitch patterns to convey lexical meaning (Yip, 2002). Mandarin Chinese, for instance, features four lexical tones: high-level, rising, dipping, and falling. In music, pitch is recorded as distinct notes, with sequences of pitches forming melodies (Krishnan et al., 2009).

Received: September 16, 2021
Funding: This research was supported by the China Postdoctoral Science Foundation General Program (2019M651881) and the Jiangsu Provincial Philosophy and Social Science Project (2018SJA0211).
Corresponding Author: LIANG Dandan, E-mail: ldd233@sina.com

1.1 The "Modularity View" versus "Shared Domain-General View" of Linguistic and Musical Pitch Processing

Pitch exhibits both commonalities and differences in its linguistic and musical manifestations, prompting divergent perspectives on pitch processing across these domains. Some research supports the "modularity view." Based on extensive brain lesion studies, Peretz and Coltheart (2003) observed that patients with deficits in one domain often showed preserved function in the other, leading them to propose that musical ability, like language, possesses dedicated programs and knowledge stores. This framework yielded the "modular model" of music processing, wherein general acoustic information is initially analyzed before being handled by specialized subsystems for musical pitch. Chen et al. (2016) and Chen et al. (2018) found no correlation between tonal and musical pitch discrimination performance among tonal language speakers, prompting their "Split Hypothesis" —the proposal that tonal and non-tonal pitch features are represented separately and activate distinct neural networks during processing.

Conversely, more studies endorse the "shared domain-general view." Building on connections between linguistic and musical abilities, Patel (2008) suggested that these domains may share processing resources. Bradley (2013) argued that tonal pitch characteristics (direction, slope, height) map onto musical pitch features (contour, interval, key) at the acoustic level due to their physical commonalities. Numerous musician studies indicate that musical experience and training enhance tonal perception abilities (Schön et al., 2004; Marques et al., 2007; Lee & Hung, 2008; Marie et al., 2011). Besson et al. (2011) consequently proposed a training transfer effect, positing that shared neural mechanisms between language and music allow long-term musical experience to influence both common acoustic processing and domain-specific language perception. Patel (2012) further argued that musical training induces neuroplastic changes that affect speech processing. These studies primarily approach the issue from a music-to-language perspective.

Bidelman et al. (2013) theoretically proposed that neuroplastic changes induced by language/music processing are bidirectional, suggesting that tonal language speakers should also demonstrate advantages in musical pitch perception. Empirical studies support this view, revealing that tonal language speakers outperform non-tonal language speakers in musical pitch perception at both behavioral and cortical levels, indicating that advantages in one domain transfer to another (Pfordresher & Brown, 2009; Giuliano et al., 2011; Bidelman et al., 2013; Weidema et al., 2016). However, these cross-domain studies have primarily focused on domain-general pitch characteristics—namely, fundamental frequency changes at the acoustic level—without addressing the domain-specific perceptual properties of lexical tone as a linguistic element.

1.2 The Influence of Tonal Categorical Perception on Musical Pitch Perception

Phonemes represent the smallest sound units capable of distinguishing meaning in language, categorized according to their social attributes in speech. Phonemes divide into segmental and suprasegmental types: the former distinguishes different phonemes based on sound quality (e.g., different vowels and consonants), while the latter distinguishes phonemes based on properties beyond sound quality, such as pitch and duration (e.g., lexical tones). Liberman et al. (1957) discovered that English speakers exhibit categorical perception (CP) for consonant phonemes. Categorical perception refers to listeners' tendency, shaped by their native phonology, to identify multiple distinct sounds along a phonetic continuum as belonging to a limited set of categories (Liberman et al., 1957). Inspired by this work, subsequent research on the suprasegmental phoneme of tone revealed that tonal language speakers perceive different tonal continua categorically, whereas non-tonal language speakers demonstrate continuous perception (Wang, 1976; Hallé et al., 2004; Xu et al., 2006; Peng et al., 2010; Shen & Froud, 2016). Continuous perception relies entirely on the physical properties of tones, preventing listeners from perceiving acoustically equidistant stimuli as discrete, limited phonemes (Xu et al., 2006).

Thus, categorical perception of tones requires listeners to ignore most pitch differences between stimuli and perceive them as a limited number of phonemes, whereas musical pitch perception demands sensitivity to acoustic pitch differences for accurate note identification. This creates a fundamental difference between tonal and musical perception. Combined with the shared domain-general view, this presents a paradox: if language and music share processing mechanisms, tonal language speakers' categorical perception patterns might transfer to musical pitch perception, causing them to perceive acoustically distinct pitches as limited categories and thereby inhibiting fine-grained musical pitch perception.

Previous research has explored this issue from three angles, each with limitations. First, studies on cross-domain effects of categorical perception patterns found that tonal language speakers exhibit quasi-categorical perception for non-linguistic stimuli, suggesting that tonal categorical perception does transfer across domains (Hallé et al., 2004; Xu et al., 2006; Peng et al., 2010). In other words, listeners may transfer the pattern of "perceiving different pitch stimuli as a limited number of phonemes" from language to other domains, inhibiting fine pitch perception. However, these studies did not explicitly address the impact of categorical perception on musical pitch perception. Second, research on native tone' s cross-domain inhibitory effects noted that native tonal experience might inhibit musical pitch processing (Bent et al., 2006; Peretz et al., 2011; Tong et al., 2018). For instance, Peretz et al. (2011) found that Mandarin speakers performed worse than non-tonal language speakers in perceiving small descending pitch intervals, attributing this to Mandarin' s falling tone typically involving large fundamental frequency ranges that reduce sensitivity to descending pitches. Yet these studies did not clarify the relationship between this inhibitory effect and categorical perception. Third, only three studies have directly examined how tonal categorical perception affects musical pitch perception—Chang et al. (2016), Weidema et al. (2016), and Bidelman and Walker (2017)—with contradictory conclusions. The first two found that tonal categories might affect sensitivity to musical pitch changes, supporting transfer effects, while the latter found that Chinese speakers showed higher categorical perception for language than music at both behavioral and neural levels, suggesting domain specificity. Notably, all three studies failed to adequately match materials across domains. Chang et al. (2016) used three-note musical discrimination tasks that differed substantially from typical categorical perception paradigms, with small pitch intervals making it unclear whether poor Chinese group performance stemmed from cross-domain inhibition or excessive task difficulty. Weidema et al. (2016) conducted tone identification tasks with Mandarin speakers, Dutch non-musicians, and Dutch musicians, finding that Mandarin speakers and Dutch musicians outperformed Dutch non-musicians in perceiving rising and falling categories, but lacked discrimination tasks. Moreover, their "categories" referred to domain-general classes (e.g., rising, falling, level tones present in both language and music), making it impossible to determine whether musical categorical perception originated from language transfer. Bidelman and Walker (2017) used mismatched linguistic and musical materials: linguistic stimuli were vowel continua manipulating the first formant rather than fundamental frequency, while musical stimuli were major-to-minor third continua manipulating fundamental frequency.

1.3 The Influence of Native Tonal Complexity Differences on Cross-Domain Pitch Perception

Tonal complexity varies across tonal languages and dialects. Mandarin Chinese has four tones, while some languages and dialects feature more complex systems: Thai has five tones, and Cantonese, Lao, and Vietnamese each have six. These languages contain multiple level, rising, or falling tones that differ only in pitch height. Some studies suggest that speakers of tonally complex languages possess finer pitch perception abilities (Chen et al., 2012; Zheng et al., 2012; Zheng et al., 2014). Zheng et al. (2014) used ERP to demonstrate that Cantonese speakers, compared to Mandarin speakers, showed stronger late mismatch negativity for between-category than within-category stimuli in non-linguistic contexts (synthetic sawtooth waves), while Mandarin speakers showed no such effect. This implies that Cantonese speakers transfer their refined pitch perception abilities to non-linguistic perception, suggesting that tonal system complexity may exert cross-domain effects on fine-grained pitch perception. However, no research to date has confirmed the influence of native tonal complexity on musical pitch perception.

In summary, while cross-domain studies on tonal language speakers have revealed facilitative effects of native experience on musical pitch processing, most have focused on domain-general pitch transfer without considering domain-specific linguistic factors—namely, the role of categorical perception patterns and the impact of different tonal language system complexities. This study applies phonetic categorical perception continua to musical pitch perception, matching pitch characteristics across speech and musical stimuli. Using native speakers of Mandarin (relatively simple tonal categories), Vietnamese (relatively complex tonal categories), and Russian (non-tonal control), we address two questions: (1) Can native tonal categorical perception patterns transfer to musical pitch perception? (2) Can differences in native tonal system complexity produce cross-domain effects on fine-grained musical pitch perception?

For the first question, we hypothesized that tonal language speakers' categorical perception patterns would activate and transfer when perceiving musical pitches matching their native tonal pitch ranges. For the second question, we selected Vietnamese as a representative complex tonal language. Standard Vietnamese has six tonal categories (Table 1) with richer acoustic features—including contour complexity, pitch height at onset and offset, and tonal values—than Mandarin Chinese. We hypothesized that tonal system complexity differences would cross-modally influence musical pitch perception, with Vietnamese speakers demonstrating superior fine-grained musical pitch perception compared to Chinese speakers. Investigating these questions clarifies the relationship between linguistic and musical pitch perception and illuminates the modularity versus shared domain-general debate from the perspective of tonal language specificity.

Table 1 Tonal categories and tonal values for Mandarin Chinese and Standard Vietnamese
Note: Vietnamese data adapted from Xian (2016).

Participants

Ninety individuals participated in the experiment: 30 native Mandarin speakers (hereafter "Chinese group"; 14 male, 16 female; mean age 21.9 years), 30 native Vietnamese speakers (hereafter "Vietnamese group"; 13 male, 17 female; mean age 22.1 years), and 30 native Russian speakers (hereafter "Russian group"; 15 male, 15 female; mean age 21.2 years)[1]. Both the Vietnamese and Russian groups had received less than one month of Chinese language training. No participants had received professional vocal or instrumental training from institutions; only a few had received amateur training for less than three years. All participants spent their childhoods in their home countries without residence abroad exceeding six months. Medical examinations confirmed normal hearing and absence of speech, language, or neurological disorders. All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). Musical ability was assessed using the Mini-PROMS (Zentner & Strauss, 2017), and amusia was evaluated using the Montreal Battery of Evaluation of Amusia (MBEA; Peretz, Champod et al., 2003). No participants showed abnormal musical ability or amusia. One-way ANOVAs on group assessment results revealed no significant differences: Mini-PROMS, $F(2, 89) = 0.62$, $p = 0.540$; MBEA, $F(2, 89) = 1.51$, $p = 0.227$. Participants received compensation for their participation.

[1] To avoid potential dialect interference, Chinese participants were native speakers of Northern dialects or Southern Mandarin speakers who used Mandarin daily at Level 2A or above. Northern dialects contain no more than six tones. Vietnamese participants hailed from Hanoi and surrounding areas and used Standard Vietnamese daily.

Materials

The experiment employed categorical perception paradigms comprising identification and discrimination tasks. Materials consisted of speech stimuli and musical stimuli.

Original speech samples comprised the syllable [i] without contextual cues, recorded from a male Mandarin speaker with Level 2A certification using Audition software at a 44.1 kHz sampling rate. Using Praat software, we constructed a continuum from [i] (Mandarin Yinping/Vietnamese Transverse) to [i] (Mandarin Yangping/Vietnamese Acute). The Yinping/Transverse tone' s onset and offset pitch values were set at 225 Hz. While maintaining the offset value constant, we systematically lowered the onset value to 175 Hz, creating a 50 Hz range spanning nine stimuli. During synthesis, onset pitch values were converted from Hertz (Hz) to Equivalent Rectangular Bandwidth (ERB) scale—a psychoacoustic standard that accurately reflects perceived frequency changes (Greenwood, 1961). Individual stimulus duration was 500 ms, approximating monosyllabic durations in Mandarin and Vietnamese (Guo, 1993; Phan, 2008). Using Praat's high-resolution pitch extraction algorithm, we extracted pitch contours from the nine speech stimuli and synthesized homologous continuous sine waves to simulate musical stimuli, ensuring identical pitch contours, amplitude, and duration between speech and musical counterparts. Figure 1 illustrates the pitch patterns for both continua.

Table 2 presents onset and offset pitch parameters for each stimulus and their actual perceptual outcomes within each domain. For Stimulus 9, onset and offset pitches were 175 Hz and 225 Hz respectively, perceived linguistically as Mandarin Yangping and Vietnamese Acute, and musically approximating a glissando from great staff F (notated as 4) to small staff A (notated as 7). Waveform and spectrogram examples for Stimulus 9 appear in Figure 2.

Table 2 Onset pitch parameters for continuum stimuli
[Table content preserved with step size in Hz and ERB, tonal categories]

Procedure

The experiment followed classic categorical perception research paradigms, comprising identification and discrimination tasks completed on computers using E-prime 3.0 software.

The identification task employed an ABX procedure. Participants heard three stimuli per trial and, after the third target stimulus, pressed a key to indicate whether its tone matched the first stimulus (Speech Stimulus 1: Yinping/Transverse; Musical Stimulus 1: A) by pressing "A" or the second stimulus (Speech Stimulus 9: Yangping/Acute; Musical Stimulus 9: F→A) by pressing "L". Each of the nine speech stimuli appeared 10 times, as did each musical stimulus, totaling 180 trials. Each trial began with a 500 ms visual cue (white cross) on a black screen, followed by sequential presentation of Stimulus 1, Stimulus 9, and the target stimulus with 500 ms inter-stimulus intervals.

The discrimination task used an AX procedure. Participants heard stimulus pairs and judged whether the two stimuli were identical (press "A") or different (press "L"). Stimulus pairs fell into two categories: different pairs with two-step separations (e.g., "1-3" from the continuum, plus "2-4", "3-5", "4-6", "5-7", "6-8", "7-9"), each repeated 10 times with both orders (e.g., "1-3" and "3-1") presented five times each; and identical pairs from "1-1" to "7-7". With 500 ms inter-stimulus intervals, each stimulus pair repeated 10 times. Both speech and musical stimuli comprised 160 stimulus pairs, totaling 320 trials. Trials began with a 500 ms visual cue (white cross) before stimulus pair presentation.

Trials were presented in pseudorandom order to prevent consecutive identical trials. Stimulus presentation level was 80 dB. After participant response or 2500 ms without response, a 500 ms interval preceded the next trial. Half the participants completed language tasks first, then music tasks; the other half completed them in reverse order. Practice trials using different stimuli preceded each formal task, continuing until participants fully understood the requirements.

Data Analysis

Following Xu et al. (2006), Peng et al. (2010), and Jiang et al. (2012), identification task evaluation used two metrics: identification regression curve slope and category boundary width. Discrimination task evaluation employed three metrics: within-category discrimination rate, between-category discrimination rate, and peakness.

Identification task results underwent logistic regression analysis to obtain the regression equation:

$$P1 = \frac{1}{1 + e^{-(b0+b1x)}}$$

where *P1* represents the probability of participants identifying experimental stimuli as Yinping/Transverse/A, x represents the nine stimuli, *b0* is the intercept, and *b1* is the slope. Larger absolute *b1* values indicate higher categorization. Category boundary width was defined as the linear distance between *P1* = 0.25 and *P1* = 0.75 on the identification curve; smaller widths indicate higher categorization.

For discrimination tasks, all stimulus pairs were divided into seven stimulus-pair groups, each containing four pair types. For example, group "1-3" included "1-3", "3-1", "1-1", and "3-3". Data were analyzed by stimulus-pair group. Discrimination rate *P(i,j)* was calculated as:

$$P(i, j) = P_S \cdot P(S) + P_D \cdot P(D)$$

where S represents the number of identical stimulus pairs, D represents the number of different pairs, *P(S)* is the proportion of identical pairs to total pairs, *P(D)* is the proportion of different pairs, *P_S* is the ratio of "same" responses to all identical pairs, and *P_D* is the ratio of "different" responses to all different pairs. Correct rates were calculated for each stimulus-pair group. Between-category discrimination rate was the mean discrimination accuracy for pairs crossing the identification curve' s category boundary (the stimulus value corresponding to 0.5 identification rate). For instance, if a participant' s category boundary was 5.1, their between-category discrimination rate was the average accuracy for crossing pairs "4-6" and "5-7". Within-category discrimination rate was the mean accuracy for the remaining seven stimulus-pair groups. Peakness was the difference between between-category and within-category discrimination rates. Higher categorization corresponds to lower within-category discrimination rates but higher between-category discrimination rates and peakness values.

3.1 Identification Task

Figure 3 displays identification rates for language and music stimuli across the three groups. Chinese and Vietnamese groups showed similar identification rates for speech and musical stimuli, with near-100% accuracy at continuum endpoints and abrupt drops near the middle, conforming to categorical perception characteristics. The Russian group showed substantially different patterns for both speech and music stimuli, approximating continuous perception.

Table 3 presents identification curve slopes and category boundary widths:

Table 3 Identification task metrics by group (n = 90)
[Table content preserved with slope and boundary width values, standard deviations in parentheses]

A 3 (group: Vietnamese, Chinese, Russian) × 2 (stimulus type: speech, music) repeated-measures ANOVA on identification curve slope revealed a significant main effect of stimulus type, $F(1, 87) = 6.19$, $p = 0.015$, ^{2} = 0.07. Post-hoc analysis showed speech stimulus identification slopes were significantly shallower than music stimulus slopes, $p = 0.015$, 95% CI = [-0.03, 0.27]. The main effect of group was significant, $F(2, 87) = 58.09$, $p < 0.001$, ^{2} = 0.57. The stimulus type × group interaction was significant, $F(2, 87) = 3.29$, $p = 0.042$. Simple effects analysis of group within stimulus type showed that for speech stimuli, the Chinese group performed significantly better than the Vietnamese group, $p = 0.002$, 95% CI = [0.22, 0.90]; the Chinese group outperformed the Russian group, $p < 0.001$, 95% CI = [-1.39, -0.71]; and the Vietnamese group outperformed the Russian group, $p < 0.001$, 95% CI = [-1.94, -1.26]. For musical stimuli, no significant difference emerged between Chinese and Vietnamese groups, $p = 0.186$; the Chinese group outperformed the Russian group, $p < 0.001$, 95% CI = [-1.72, -1.03]; and the Vietnamese group outperformed the Russian group, $p < 0.001$, 95% CI = [-1.95, -1.26].

For boundary width, the main effect of stimulus type was non-significant, $F(1, 87) = 0.87$, $p = 0.352$. The main effect of group was significant, $F(2, 87) = 3.40$, $p = 0.038$, ^{2} = 0.07. Post-hoc analysis revealed no significant difference between Chinese and Vietnamese groups, $p = 0.980$; the Chinese group showed significantly greater boundary width than the Russian group, $p = 0.027$, 95% CI = [-33.01, -2.21]; and the Vietnamese group showed significantly greater boundary width than the Russian group, $p = 0.026$, 95% CI = [-32.82, -2.01]. The stimulus type × group interaction was non-significant, $F(2, 87) = 0.83$, $p = 0.352$. These results indicate that Chinese and Vietnamese groups performed similarly and differed substantially from the Russian group.

3.2 Discrimination Task

Figure 4 illustrates discrimination results for language and music stimulus pairs across groups, with discrimination metrics presented in Table 4:

Figure 4 Discrimination task results by group
Table 4 Discrimination task metrics by group (n = 90)
[Table content preserved with within-category, between-category, and peakness values, standard deviations in parentheses]

A 3 (group: Vietnamese, Russian, Chinese) × 2 (stimulus type: speech, music) repeated-measures ANOVA on within-category discrimination rate showed no significant main effect of stimulus type, $F(1, 87) = 0.001$, $p = 0.972$. The main effect of group was significant, $F(2, 87) = 4.95$, $p = 0.009$, ^{2} = 0.10. The stimulus type × group interaction was non-significant, $F(2, 87) = 0.92$, $p = 0.404$. Post-hoc analysis revealed that the Vietnamese group performed significantly better than the Russian group, p = 0.017, 95% CI = [0.01, 0.11]; the Chinese group outperformed the Russian group, p = 0.004, 95% CI = [0.02, 0.12]; while Chinese and Vietnamese groups did not differ significantly, p = 0.615.

For between-category discrimination rate, the main effect of stimulus type was non-significant, $F(1, 87) = 0.36$, $p = 0.550$. The main effect of group was significant, $F(2, 87) = 8.27$, $p = 0.001$, $^2 = 0.16$. The interaction was non-significant, $F(2, 87) = 0.87$, $p = 0.422$. Post-hoc analysis showed the Vietnamese group outperformed the Russian group, p = 0.009, 95% CI = [0.02, 0.15]; the Chinese group outperformed the Russian group, $p < 0.001$, 95% CI = [0.07, 0.20]; while Chinese and Vietnamese groups did not differ significantly, p = 0.184.

For peakness, the main effect of stimulus type approached significance, $F(1, 87) = 3.52$, $p = 0.064$. The main effect of group was significant, $F(2, 87) = 8.80$, $p < 0.001$, $^2 = 0.17$. Post-hoc analysis revealed that the Vietnamese group outperformed the Russian group, p = 0.002, 95% CI = [0.04, 0.17]; the Chinese group outperformed the Russian group, $p < 0.001$, 95% CI = [0.07, 0.20]; while Chinese and Vietnamese groups did not differ significantly, p = 0.487. The interaction was non-significant, $F(2, 87) = 0.87$, $p = 0.422$. These results indicate that the two tonal language groups performed similarly to each other and differed substantially from the Russian group.

4.1 Cross-Domain Transfer of Tonal Categorical Perception Patterns

The first research question addressed whether native tonal categorical perception patterns transfer to musical pitch perception. Results demonstrated that both Chinese and Vietnamese groups exhibited categorical perception for linguistic and musical pitches, with no stimulus-type differences in category boundary width, within-category discrimination rate, between-category discrimination rate, or discrimination peak. This indicates that native categorical perception patterns transfer to musical pitch perception, consistent with our hypothesis. The Russian group showed no abrupt shifts in identification curves and displayed relatively flat discrimination curves with multiple peaks, indicating consistent continuous perception based on acoustic properties for both speech and music stimuli that differed significantly from tonal language groups on all five metrics.

Previous language-to-music transfer studies have primarily used two-note or musical fragment discrimination tasks (Pfordresher & Brown, 2009; Giuliano et al., 2011; Hutka et al., 2015), finding that tonal language speakers outperform non-tonal language speakers due to enhanced sensitivity from tonal contour processing experience (Giuliano et al., 2011). In contrast, the present study examined transfer of categorical perception—a phoneme-level tonal processing factor—and found that both tonal language groups displayed patterns for musical stimuli similar to those for speech stimuli, confirming that tonal categorical perception transfers to musical pitch perception. These results support the shared domain-general view.

Our findings align with Chang et al. (2016) and Weidema et al. (2016) but differ from Bidelman and Walker (2017), possibly because the latter failed to match linguistic and musical stimuli. Typically, linguistic tones and musical notes belong to different representational systems: tones are categorized while musical pitches are represented discretely (Bradley, 2013). Musicians cannot directly transfer interval knowledge to tonal perception, nor can tonal language speakers directly transfer categorical perception to musical melody perception. Transfer effects only occur when specific pitch category experience from one domain matches the pitch information requiring processing in another (Asaridou & McQueen, 2013). Bidelman and Walker (2017) used vowel continua for linguistic stimuli and major-to-minor third continua for musical stimuli—non-homogeneous materials with mismatched pitch information that prevented activation of speech categorical perception patterns during musical pitch discrimination.

4.2 Influence of Native Tonal System Complexity Differences on Musical Pitch Perception

The second research question examined whether native tonal system complexity differences affect fine-grained musical pitch perception. Fine-grained pitch perception is typically assessed via pitch discrimination tasks (Pfordresher & Brown, 2009; Giuliano et al., 2011; Hutka et al., 2015). Our discrimination task measured participants' ability to differentiate speech and musical tones of varying pitch, reflecting sensitivity to fine pitch differences. However, results showed no significant differences between Vietnamese and Chinese groups in within-category or between-category discrimination rates for either linguistic or musical conditions, indicating that complex native tonal systems do not facilitate cross-domain fine-grained musical pitch perception. This finding failed to support our hypothesis.

Two factors may explain this result. First, from a linguistic perspective, our materials comprised a Mandarin Yinping-Yangping/Vietnamese Transverse-Acute continuum. Vietnamese participants activated Vietnamese tonal representations that were essentially equivalent to the Mandarin representations activated by Chinese participants, yielding no significant between-group differences in between-category discrimination rates. Compared to Mandarin, Vietnamese tonal complexity manifests in categories like Huyền and Ngã tones not included in our materials, preventing detection of the refined tonal processing abilities conferred by Vietnamese' s complex system. Peng et al. (2010) and Zheng et al. (2012) reported similar findings when comparing Mandarin to the more complex Cantonese system: behavioral studies found no acoustic-level differences between Cantonese and Mandarin speakers in Mandarin pitch perception. However, Zheng et al.' s (2012) ERP study revealed that Cantonese speakers' fine-grained pitch perception advantages transferred to Mandarin tonal perception at the cortical level, effects invisible at the behavioral level. Future neurophysiological research should further investigate cross-domain effects of tonal complexity differences on musical pitch perception.

Second, from a musical perspective, both tonal language groups showed significantly higher between-category discrimination rates for musical stimuli than the Russian group, demonstrating significant cross-domain benefits of tonal language experience for musical pitch discrimination. This language-to-music direction supports the shared domain-general view and Besson et al.' s (2011) training transfer hypothesis. However, according to the OPERA hypothesis (Patel, 2012), musical experience more effectively promotes fine-grained pitch perception than language experience because music perception requires continuous pitch adjustment and monitoring with higher encoding precision demands. Empirical studies support this: tonal language speakers are outperformed by musicians in discriminating non-native tones based on acoustic information (Cooper & Wang, 2012; Chang et al., 2016). Thus, despite Vietnamese' s greater tonal complexity than Mandarin, its tonal inventory and pitch granularity remain simpler than musical melodies, preventing Vietnamese speakers from demonstrating cross-domain facilitative effects at the behavioral level.

This study examined cross-domain transfer of linguistic factors affecting tonal processing to musical pitch perception. Findings demonstrate that tonal categorical perception patterns transfer to musical pitch perception, while complex native tonal systems do not enhance cross-domain fine-grained musical pitch perception. These results support the shared domain-general view regarding language' s influence on musical processing.

References

[All references preserved exactly as in original, including Chinese references with English translations]

Note: Figure translations are in progress. See original paper for figures.