Research Article

Effect of Musical Aptitude on the Perception of English Vowels: An Eye-Tracking Investigation Among Native Mandarin Speakers

Jiayu Liang,[a] Hao Zhang,[a] Wen Ma,[a] and Hongwei Ding[b]

[a] Center for Clinical Neurolinguistics, School of Foreign Languages and Literature, Shandong University, Jinan, China [b] Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, China

ARTICLE INFO

Article History:
Received December 25, 2024
Revision received April 26, 2025
Accepted May 10, 2025

Editor-in-Chief: Rachael Frush Holt
Editor: Gavin M. Bidelman

https://doi.org/10.1044/2025_JSLHR-24-00916

ABSTRACT

Purpose: Previous research has suggested that individuals' higher musical aptitude enhances their speech perception in terms of pitch and temporal features. However, it remains unclear whether this cross-domain transfer could extend to the perception of second language (L2) vowels. The primary aim of this study is to investigate how musical aptitude influences the categorical perception of English vowels by native Mandarin speakers.

Method: Sixty Mandarin speakers were assigned into a high aptitude (HA) group and a low aptitude (LA) group based on the median of their musical aptitude test scores. Each participant completed a visual-world eye-tracking experiment on the categorical perception of English vowels, which included the acoustically less salient /ɛ/–/æ/ contrast and the more salient /i/–/eɪ/ contrast. Statistical analyses were conducted on both behavioral and eye-tracking data to compare vowel categorization across different groups and vowel contrasts.

Results: Overall, the HA group significantly outperformed the LA group in vowel categorization. In addition, the Group × Contrast interaction on boundary width and correlational results on the eye-tracking parameter showed more robust effect of musical aptitude observed for the acoustically less salient /ɛ/–/æ/ contrast.

Conclusion: Mandarin speakers with relatively higher musical aptitude tend to show more refined categorization of L2 English vowels, with this cross-domain transfer effect modulated by acoustic salience.

The potential relationship between music and language is a topic that is ripe for investigation. A considerable amount of evidence deriving from neural and behavioral research has demonstrated the cross-domain transfer of musical training–induced benefits to vowel perception (Bidelman & Alain, 2015; Bidelman & Krishnan, 2010; Bidelman et al., 2014; Butera, 2015; Choi et al., 2024; Cooper et al., 2016; Elmer et al., 2017; Hutka et al., 2015; Kühnis et al., 2013; Marie et al., 2011; Sadakata & Sekiyama, 2011; Z. Zhang et al., 2023). These cross-domain benefits appear to be modulated by acoustic salience (Choi et al., 2024; Ong et al., 2020; Sadakata & Sekiyama, 2011; Toh et al., 2023). However, several more recent studies have suggested that preexisting factors—such as musical aptitude—could play a pivotal role in determining far-transfer effects (i.e., the generalization of knowledge and skills obtained from music training to a substantially different domain such as speech processing, in stark contrast to near-transfer effects that are generalized within the music domain) post musical training, challenging the overestimation of musical experience per se in enhancing speech perception (Jansen et al., 2023; Kragness et al., 2021; Schellenberg, 2015; Schellenberg & Lima, 2024; Swaminathan & Schellenberg, 2017). Despite these implications, little research has examined how musical aptitude affects vowel perception and how acoustic salience might modulate this cross-domain transfer. To address

Correspondence to Hao Zhang: hao.zhang0099@sdu.edu.cn. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

these gaps, the overarching goal of this study was to investigate the effect of preexisting musical aptitude on the perception of English vowels with varying acoustic salience among Mandarin-speaking adults, adopting categorical perception (CP) and visual world paradigms (VWPs).

The Selective Effect of Musical Aptitude on Second Language Vowel Perception

As Jansen et al. (2023) pointed out, it is well-established that musical abilities—whether acquired through formal training or stemming from innate musical aptitude—are closely linked to speech prosody perception. Accumulating studies consistently demonstrated that individuals with stronger musical abilities tend to be more sensitive to pitch-related features, particularly fundamental frequency (*F*0), which is central to the perception of tone and intonation (Bowles et al., 2016; Choi, 2022; Cui & Kuang, 2019; Yao et al., 2022; K. Zhang et al., 2023; Zhu et al., 2021). This advantage is often attributed to enhanced spectral processing skills that transfer across domains (Musso et al., 2020). Importantly, spectral processing is not limited to pitch (*F*0); it also plays a key role in formant perception (Kempe et al., 2015; Musso et al., 2020; C. Zhang et al., 2017). *F*0, the acoustic correlate of pitch, is determined by the frequency of the first harmonic and the distance between neighboring harmonics of the speech spectrum (Ladefoged & Johnson, 2006), suggesting that pitch processing depends on frequency/spectral analysis (C. Zhang et al., 2017). Similarly, the decoding of vowels also relies on frequency/spectral processing. For example, vowels are acoustically characterized by the frequency location of spectral peaks (i.e., formants) in the speech spectrum, with the first two formants (F1 and F2) serving as the primary cue for distinguishing most of the vowels (Ladefoged & Johnson, 2006; Peterson & Barney, 1952). Neuroimaging evidence further supports this association: Spectral cues such as *F*0 and formants are both related to increased right-hemisphere activation (Kühnis et al., 2013), and musical pitch sensitivity served as the basis for the processing of first and second formant frequencies (Choi et al., 2024; Loui et al., 2011). Taken together, these findings point to a robust association between music and vowel perception, likely rooted in their shared reliance on the processing of spectral cues (Kühnis et al., 2013).

Most existing research on the link between music and vowel perception has focused on the effects of musical training. Within this body of work, findings have been drawn from studies examining native and second language (L2) vowel perception. For native vowel perception, a number of studies have reported that musicians outperform nonmusicians (Bidelman & Alain, 2015; Bidelman & Krishnan, 2010; Bidelman et al., 2014; Butera, 2015; Elmer et al., 2017; Hutka et al., 2015; Kühnis et al., 2013; Z. Zhang et al., 2023). However, not all findings are consistent. For example, Yashaswini and Maruthy (2020) found no effect of musical training on the CP of the native vowel contrast /u/−/a/, although an effect was observed for consonants. Similar patterns have been observed in studies on L2 vowel perception. Musicians have generally shown an advantage over nonmusicians (Choi et al., 2024; Cooper et al., 2016; Marie et al., 2011), though findings remain mixed. Sadakata and Sekiyama (2011), for instance, observed a musician advantage for discriminating spectral cues, but this advantage was not consistent across all vowel contrasts. Inconsistencies in the literature may partly reflect the limitations of relying on self-reported musical experience. Swaminathan and Schellenberg (2017) found that rhythm perception, rather than formal musical training, was a better predictor of nonnative phoneme discrimination performance. These findings raise the possibility that musical aptitude—rather than training per se—may be the more foundational factor underlying cross-domain transfer.

Musical aptitude refers to an individual's natural potential in reaping musical achievement regardless of training experience (Gordon, 1989), which is directly estimated via measurements of inherent ability to perceive and produce musical-related acoustic elements, such as pitch, rhythm, and harmony (Jansen et al., 2023; Schellenberg & Lima, 2024). Various standardized tools have been developed to measure musical aptitude, including Advanced Measure of Musical Audiation (Gordon, 2007), Musical Ear Test (Wallentin et al., 2010), Profile of Music Perception Skills (Law & Zentner, 2012), Swedish Musical Discrimination Test (Ullén et al., 2014), and Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2008). While these assessments vary in scope and emphasis, they consistently rely on closed-set, forced-choice discrimination of melodic, rhythmic, or pitch-based stimuli to evaluate perceptual musical aptitude. As Schellenberg (2015) pointed out, musical aptitude could be considered as a more foundational factor contributing to both involvement and achievement in the music domain. Moreover, it is proposed that musical aptitude has a close relationship with nonmusical abilities of basic auditory acuities (Baldé et al., 2025; Lehmann et al., 2015; Oikkonen et al., 2015; Schneider et al., 2002; Shahin et al., 2007; Zheng et al., 2022) and language skills (Choi, 2022; Jansen et al., 2023; Mankel et al., 2020; Mankel & Bidelman, 2018; Qin et al., 2021; Schellenberg, 2015; Schellenberg & Lima, 2024; Strait et al., 2011; C. Zhang et al., 2017). Recent studies suggest that musical aptitude, rather than formal musical training, serves as a more general impulse underlying the cross-domain transfer to enhancing speech perception (Schellenberg & Lima, 2024).

The link between musical ability and vowel perception has been predominately examined in terms of acquired musical experience, with comparatively limited investigation on innate musical abilities. In particular, current findings remain inconsistent regarding the role of musical aptitude in influencing nonnative vowel perception. For example, although Kempe et al. (2015) reported better discrimination of nonnative vowel contrasts of /i/–/y/ and /iː/–/yː/ among participants with higher musical aptitude, several more recent studies failed to observe significant correlations between overall scores of musical ability and perceptual outcomes of nonnative vowels (Ghaffarvand Mokari & Werner, 2018; Smit et al., 2022). Crucially, the extant cross-linguistic observations mainly focused on Western populations, with Mandarin–English bilinguals underestimated. The perception of L2 English vowels among Mandarin-speaking learners merit further investigation given phonological transfer challenges and the overall elevated pitch sensitivity (i.e., a perceptual dimension where musical aptitude may potentially exert privileged influence). In addition, previous research exhibited over-reliance on perceptual responses of isolated vowels or nonword syllables, with relatively limited examination on real-time processing dynamics of monosyllabic words. To bridge these research gaps, this study adopted ecologically valid stimulus and time-sensitive measurement for the investigation of musical aptitude's role in L2 vowel perception among Mandarin speakers.

Previous research indicated that the influence of music-related factors on speech perception varies across different types of speech sounds, including tones and consonants (Choi et al., 2024; Ma et al., 2024; Yao et al., 2022), as well as vowels (Jekiel & Malarski, 2021). Moreover, the cross-domain transfer effects may be modulated by acoustic salience—defined as the degree of acoustic or perceptual distinctiveness between phonetic categories (Narayan et al., 2010; see also Choi et al., 2024; Ong et al., 2020; Sadakata & Sekiyama, 2011; Toh et al., 2023). Additionally, the way in which acoustic salience modulates the musical advantage has produced mixed findings. On one hand, several studies have reported stronger effects of musical training in conditions involving acoustically less salient contrasts, where listeners with musical training outperformed nonmusicians (Choi et al., 2024; Toh et al., 2023; Ong et al., 2020). On the other hand, Sadakata and Sekiyama (2011) found greater group differences in conditions with more salient contrasts, where musicians showed significantly higher accuracy than nonmusicians. Despite these mixed findings, relatively little attention has been given to how acoustic salience may shape the influence of musical aptitude more broadly. For instance, Jekiel and Malarski (2021) found that Polish learners with stronger rhythmic memory produced the English vowel /æ/ more natively than other vowels, indicating a selective advantage tied to certain acoustic properties. More importantly, previous research that found the positive effect of musical aptitude did not include vowel contrasts with varying acoustic salience (Kempe et al., 2015; C. Zhang et al., 2017). To address this gap and systematically investigate how musical aptitude interacts with acoustic salience, the present study examines two vowel contrasts that differ in acoustic distinctiveness—one with higher salience and one with lower salience—allowing for a more precise understanding of whether and how musical aptitude differentially enhances L2 vowel perception based on acoustic salience.

Assessing L2 Vowel Categorization With VWP

The categorical versus continuous accounts in vowel perception remain a persisting theoretical issue in speech processing. CP refers to the abstraction of phonetic categories, which is characterized by the perceptual mapping of infinite acoustic variations onto a finite set of phonemic labels (Harnad, 1987; Liberman et al., 1957; Y. Zhang, 2016). While plosive consonants exhibit robust CP (Liberman et al., 1957), seminal work by Fry et al. (1962) posited continuous pattern in vowel perception. However, more contemporary research proposed a categorical-continuous spectrum of vowel perception, with a range of variables modulating categorical positions on the spectrum (Chen et al., 2019). Several principal factors have been revealed, including linguistic contexts (Repp et al., 1979), task demands (Pisoni, 1975), and language experiences (Stevens et al., 1969; H. Zhang et al., 2016). It is noteworthy that Chen et al. (2019) demonstrated enhanced degree of CP for diphthongs relative to monophthongs in native Mandarin speakers, as acoustics of diphthongs showing much greater dynamic pattern in spectral transitions. Emerging evidence further suggests domain-general auditory enhancements through musical training (Bidelman & Alain, 2015; Bidelman & Krishnan, 2010; Bidelman et al., 2014; Butera, 2015; Elmer et al., 2017; Hutka et al., 2015; Kühnis et al., 2013) and innate musical aptitude (Mankel & Bidelman, 2018; Yashaswini & Maruthy, 2020; C. Zhang et al., 2017) might sharpen categorical boundaries in speech perception. Crucially, this body of research suggests that the CP paradigm is a feasible and robust method for investigating the transfer effects of musical effects on speech perception. Building on this line of research, the present study adopts the CP measurement to examine how individual differences in musical aptitude modulate the perception of L2 English vowels among Mandarin speakers.

The classical CP paradigm typically incorporates identification and discrimination tasks that predominantly depend on endpoint measures of behavioral responses, neglecting the real-time cognitive processes preceding perceptual decisions. While effective for assessing phonetic categorization, both tasks may limit insights into the temporal dynamics of speech processing as acoustic signals unfold. Building on the well-established methodology by McMurray and colleagues (McMurray et al., 2002, 2008, 2018), the current study adopts the eye-tracking technique with VWP. As a time-sensitive approach, this technique captures the dynamic time course of perceptual decision-making, offering a more fine-grained investigation on how speech sounds are processed over time. Following previous research (McMurray et al., 2002, 2008, 2018), a dual-modality design synergistically combines VWP with discrete labeling procedures (i.e., behavioral identification task). This task contributes to psychometric functions of identification performances that are highly indicative of the listeners' phonetic categorization abilities (Bidelman, 2015; Bidelman & Alain, 2015; Bidelman et al., 2014; Lewis & Bidelman, 2020). Therefore, the methodology permits simultaneous examination of final categorical judgments and the dynamic decision trajectories along phonetic identification (Ito et al., 2018; McHaney et al., 2021; McMurray, 2023; Qin et al., 2019; Turner, 2022). The experimental design adopted in this study provides a refined assessment to examine how musical aptitude modulates both the endpoint and incremental processes of L2 vowel perception in Mandarin speakers.

The Current Study

The primary aim of this study is to determine whether, and to what extent, musical aptitude affects the CP of L2 English vowels among native Mandarin speakers, employing the eye-tracking technique with VWP. The main hypothesis is that L2 listeners with higher musical aptitude could demonstrate enhanced categorization of English vowels, which is made on basis of previous findings that higher musical aptitude correlates with better speech perception (Kempe et al., 2015; Lehmann et al., 2015; Mankel et al., 2020; Mankel & Bidelman, 2018; Strait et al., 2011; C. Zhang et al., 2017). Moreover, prior research indicated that acoustic salience could play a pivotal role in modulating the cross-domain transfer from music to speech perception (Choi et al., 2024; Ong et al., 2020; Sadakata & Sekiyama, 2011; Toh et al., 2023). To systematically investigate this proposition, the present study employed two distinct vowel contrasts, with /ɛ/−/æ/ representing less salient contrast while /i/−/eɪ/ representing a more salient pairing. This selection of vowel contrasts leverages both psychoacoustic and cross-linguistic rationales. Perception of the /i/−/eɪ/ pairings benefited from innate acoustic prominence, with diphthongs showing greater dynamic changes in spectral variations, especially the second formant trajectory (Chen et al., 2019). Moreover, both /i/ and /eɪ/ are shared in the vowel inventories of Mandarin and English (Yang & Fox, 2017), which may further enhance their perceptual distinctiveness with improved phonological familiarity for Mandarin listeners. Therefore, the /i/−/eɪ/ continuum is expected to produce more robust CP outcomes, exhibiting less sensitivity in detecting individual differences in auditory abilities, including those related to musical aptitude (Baldé et al., 2025). In contrast, the /ɛ/−/æ/ contrast presented greater challenges to L2 English learners (Jekiel & Malarski, 2021), due to their subtle distinction in spectral separations. This is particularly true for Mandarin speakers, since the lack of a corresponding phonemic distinction in their native vowel system (Zhou et al., 2022). It follows that the perceptual ambiguity of /ɛ/−/æ/ contrast engaged greater demands on fine-grained auditory discrimination that is frequently linked to musical aptitude (Baldé et al., 2025; Lehmann et al., 2015; Oikkonen et al., 2015; Schneider et al., 2002; Shahin et al., 2007; Zheng et al., 2022). Thus, this less salient contrast may provide greater sensitivity to individual variability in perceptual skills. The employment of two vowel contrasts with different acoustic salience permits investigation on the interaction between musical aptitude and acoustic salience to shape L2 vowel categorization among Mandarin-speaking learners, thereby providing important insights into the mechanisms underpinning cross-domain transfer of musical ability to speech perception.

Method

Participants

This study recruited 60 native Mandarin-speaking undergraduates from Shandong University (M_{age} = 19.9 years, SD = 0.94), including 24 male and 36 female participants. None of them reported a history of formal musical training experience, visual impairments, auditory injuries, or psychiatric disorders. Each participant received nominal financial compensation for their participation. Before participating in the eye-tracking experiment, the subjects were measured on their musical aptitude and L2 English proficiency. Following prior studies (Chui & Qin, 2024; Cui & Kuang, 2019; Qin et al., 2021, 2022), musical aptitude of the participants was assessed using MBEA (Peretz et al., 2008). L2 English proficiency was evaluated via a suite of tools including Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012), Shipley Vocabulary Test (Shipley, 1940), and a self-reported language background questionnaire (Hui et al., 2020). Additionally, the participants were assigned into the high aptitude (HA) group (n = 30) and the low aptitude (LA) group (n = 30), based on the median of their MBEA overall scores (81.49). In general, participants in the HA group demonstrated significantly higher musical aptitude (*M* = 87.26, *SD* = 3.81) than the LA group (*M* = 71.01, *SD* = 7.44). Furthermore, there was no significant difference between the two groups in language scores or questionnaire results. Demographic information of the two groups is shown in Table 1. This study was approved by the Ethics Committee of the School of Foreign Languages, Shandong University (Ethical Approval Number: ECSFLLSDU2025–3). Informed consents were obtained from all participants.

Materials and Procedure

This study included the vowel contrasts of /ɛ/−/æ/ (i.e., a "difficult" contrast with less acoustic salience) and /i/−/eɪ/ (i.e., an "easy" contrast with more acoustic salience) to examine how acoustic salience might interact with musical aptitude in affecting CP outcomes. Specifically, the experimental materials were four pairs of English words featuring vowel contrasts of /ɛ/−/æ/ and /i/−/eɪ/, including "bet-bat," "bed-bad," "beak-bake," and "beat-bait." Following Connell et al. (2018), word frequency of these words was retrieved from the Corpus of Contemporary American English (Davies, 2008) and log-transformed. Paired-samples *t* tests showed no significant difference on the log-transformed frequencies between words for different vowel contrasts, $t(6) = -1.48$, $p = .19$. These words were recorded in plain and clear forms from a female native speaker of American English in a sound-treated booth at a sampling rate of 44.1 kHz (16 bit), which were further normalized to 75 dB SPL and 360 ms in terms of intensity level and duration, respectively. Four consonant–vowel–consonant (CVC) continua (i.e., "bet-bat," "bed-bad," "beak-bake," "beat-bait") were synthesized using the TANDEM-STRAIGHT MATLAB toolbox (Kawahara & Morise, 2011), with each consisting of 11 stimuli. The continuum synthesization was implemented according to the following steps. Initially, the *F*0 structure, aperiodicity, and spectrogram details of two endpoints (e.g., /bɛt/ and /bæt/ in the "bet-bat" continuum) were extracted as source and filter parameters. Then, a continuum from Source A (one endpoint) to Source B (the other endpoint) was morphed into 11 equidistant speech stimuli in both spectral and temporal domains using the source-filter model (Feng & Peng, 2023). For the /ɛ/−/æ/ continuum, CVC words were chosen as endpoints (i.e., "bet" and "bed" as Stimulus 1, while "bat" and "bad" as Stimulus 11). Stimulus 1 had respective F1 and F2 frequencies of 840 Hz and 2016 Hz, whereas Stimulus 11 showed an F1 frequency of 962 Hz and an F2 frequency of 1886 Hz. Similarly, CVC words were treated as endpoints in /i/−/eɪ/ continuum (i.e., "beak" and "beat" as Stimulus 1, and "bake" and "bait" as Stimulus 11). For Stimulus 1, the F1 frequency was 314 Hz, and the F2 frequency was 2879 Hz. For stimulus 11, the F1 and F2 frequencies were 425 Hz and 2463 Hz, respectively.

Adopting a four-alternative forced-choice identification task within the classical techniques of CP and VWP (cf. McMurray et al., 2002, 2008), the eye-tracking experiment was administered in two blocks. VWP (Huettig & McQueen, 2007) was administered with an auditory stimulus accompanied by a visual display of four words, including the target, a phonological competitor, and two distractors (cf. Turner, 2022). The visual displays of example trials are shown in the Appendix. To circumvent prediction and fatigue effects of the participants, a counterbalanced design was developed on the trial presentation in each block. Specifically, for half of all trials within one block, 11 auditory stimuli from the "bet-bat" or "bed-bad" continuum (/ɛ/−/æ/ vowel contrast) functioned as targets/competitors, whereas visual words featuring the /i/−/eɪ/ vowel contrast ("beak-bake" or "beat-bait") served as distractors, and vice versa for the other half of trials within the same block. That is to say, the targets/competitors were "beak-bake" or "beat-bait," while the distractors were "bet-bat" or "bed-bad." Each auditory stimulus was repeated six times, resulting in a total of

Table 1. Demographic information of two groups.

Musical and language measurements	HA (*n* = 30)	LA (*n* = 30)	Group differences	
	M (*SD*)	*M* (*SD*)	*t*	*p* value
MBEA overall score	87.26 (3.81)	71.01 (7.44)	10.64	< .0001
Age of acquisition of English	8.07 (0.94)	7.9 (1.3)	0.57	.57
Shipley Vocabulary Test	22.4 (5.82)	19.93 (8.07)	1.36	.18
LexTALE	62.67 (8.77)	58.92 (12.33)	1.36	.18
Frequency of using English (1–7)	4.0 (1.49)	3.57 (1.33)	1.19	.24
Self-rated English proficiency (1–7)	4.13 (1.07)	3.7 (0.99)	1.63	.11

Note. HA = high aptitude; LA = low aptitude; MBEA = Montreal Battery of Evaluation of Amusia; LexTALE = Lexical Test for Advanced Learners of English.

264 trials (2 continua × 11 stimuli × 6 repetitions × 2 blocks). All trials were pseudorandomized in presentation (Turner, 2022).

The experiment was designed and delivered through Experiment Builder (SR Research) that was compatible with the Eyelink Portable Duo system for the recording of online eye movements at a sampling rate of 1000 Hz. At the beginning, a 9-point calibration was administered to ensure precision in tracking. During each trial, four visual words were displayed in the corners of the screen, accompanied by a centrally positioned red circle serving as a fixation. An interstimulus interval of 2,000 ms was set to offer a preview of the location of each visual word. Afterward, the red circle transitioned to green, prompting the initiation of auditory presentation by clicking on the green circle, which ensured participants concentrated their attention on the onset of each trial (Dial et al., 2019). Then, participants were required to identify the auditory stimuli and click on the corresponding visual target. Prior to the formal trials of the eye-tracking experiment, each block included an eight-trial practice to ensure familiarity of participants with the experimental stimuli and procedures. In practice trials, any discrepancies or errors in participant responses were promptly addressed and corrected by the experimenter to maintain data integrity and accuracy. Overall, the whole experiment lasted approximately 2 hr, allowing sufficient breaks to minimize fatigue and maintain engagement for each participant.

Data Analyses

Behavioral Data

Following previous studies (Chen & Peng, 2021; Peng et al., 2010; H. Zhang et al., 2023), a Probit analysis was conducted to estimate boundary width (Finney, 1971). The boundary width was defined as the linear distance between the 25th and 75th percentiles of the curve (Peng et al., 2010). A smaller boundary width indicates a sharper transition of two phonemic categories, reflecting a clearer distinction between them (Chen & Peng, 2021). For identification data analysis, linear mixed-effects (LME) models were employed, with boundary width as the dependent variable. The fixed effects included group (HA vs. LA) and contrast (/ɛ/–/æ/ vs. /i/–/eɪ/), along with their interactions. Additionally, random intercepts and maximal slopes were incorporated as random effects (Barr et al., 2013). The significance of fixed factors was estimated with α level set at .05, and the *p* values for fixed factors were obtained using the Satterthwaite method from the lmerTest package (Kuznetsova et al., 2017). Multiple model comparisons were conducted to identify the best-fit model with the lowest Akaike information criterion. Based on the best-fit model, post hoc pairwise comparisons for significant fixed factors were carried out using the emmeans package (Lenth et al., 2018), with adjustments of false discovery rate (Benjamini & Yekutieli, 2001).

Eye-Tracking Data

For the analysis of eye-tracking data, fixation proportions for the target and competitor were aggregated into 20-ms time bins, from onset of auditory stimuli to 2,000 ms. The dependent variable for statistical analysis was the difference between empirically log-transformed fixation proportions (FP) for the target and competitor (FP difference). FP difference was widely adopted in analyses of visual-world eye-tracking data, due to the simultaneous consideration of both target and competitor activations during speech processing (Connell et al., 2018; Creel, 2014; Qin et al., 2019; Qin & Zhang, 2024). It is noteworthy that FP difference was calculated within a time window of 200–2,000 ms, excluding a 200-ms delay to account for the time that eye fixations take to reflect speech processing (Connell et al., 2018; Hallett, 1986; Qin et al., 2019; Turner, 2022). Moreover, FP difference values over 0 indicate that participants focused more on the target compared to the competitor, and vice versa (Qin et al., 2019).

FP difference was modeled using growth curve analysis (GCA; Mirman et al., 2008), a type of curvilinear regression that can fit with a third-order orthogonal polynomial to capture the linear (i.e., capturing the overall angle of a curve), quadratic (i.e., capturing a curve with a single inflection), and cubic (i.e., capturing a curve with two inflections) features of dynamic curves (Qin et al., 2019). The GCA model has been widely adopted in evaluating eye-movement patterns while speech processing unfolds (Connell et al., 2018; Ito & Knoeferle, 2023; Ito et al., 2018; Mirman et al., 2008; Qin et al., 2019). Following the manual by Dink and Ferguson (2015), FP difference was preprocessed with the EyetrackingR package in R (R Core Team, 2024). The GCA model was built for the FP difference, including linear, quadratic, and cubic time terms, with fixed effects for group (HA vs. LA), contrast (/ɛ/–/æ/ vs. /i/–/eɪ/), and their interactions. The model treated participant and item as random intercepts and the three time terms as random slopes, modeling a different curve for each participant and test item (Connell et al., 2018). Further analysis follows the same procedure of the LME analysis for the behavioral data.

Correlational Analysis

Spearman correlation was performed to examine the correlation between the measurements of perception, as well as their relationship with participants' test scores (Chen & Peng, 2021; H. Zhang et al., 2024). The measurements of perception comprised the mean boundary width and mean FP difference (Turner, 2022), while participants' test scores included MBEA scores and L2 English proficiency test scores (Shipley and LexTALE scores). This approach aimed to assess the correlation between participants' musical aptitude and vowel perception, as well as the potential effects of L2 English proficiency on processing different vowel contrasts.

Results

Behavioral Results

Figures 1 and 2 display the identification curves and boundary width of the two vowel contrasts for HA and LA groups. The best-fit LME model was specified as follows: lmer(Boundary width ~ Group × Contrasts + (1|Participant) + (1|Item), REML = FALSE). The results revealed a significant interaction between group and contrast for the boundary width, $\chi^2(1) = 93.56$, $p < .001$. Further post hoc analyses were conducted for boundary width. For boundary width, significant difference was observed for the /ɛ/−/æ/ contrast, with significantly narrower boundary width for the HA group relative to the LA group ($\beta = -5.09$, $SE = 0.40$, $t = -12.76$, $p < .0001$), indicating that the HA participants exhibited a more refined categorization for the /ɛ/−/æ/ contrast compared to the LA individuals. Additionally, boundary width of /i/−/eɪ/ was significantly narrower than that of /ɛ/−/æ/ for both HA group ($\beta = 2.56$, $SE = 0.34$, $t = 7.47$, $p = .0001$) and LA group ($\beta = 7.38$, $SE = 0.35$, $t = 21.02$, $p < .0001$), suggesting that categorization of acoustically less salient contrast of /ɛ/−/æ/ was more challenging than that of /i/−/eɪ/ contrast.

Eye-Tracking Results

Following our prediction, the HA group was expected to demonstrate enhanced vowel perception compared to the LA group, with the effect of musical aptitude being more pronounced for the /ɛ/−/æ/ contrast. Specifically, the GCA model should reveal one or more of the following effects (Qin et al., 2019): (a) a significant interaction between group and contrast on the linear time term, indicating that the FP difference increases more steeply over time (i.e., a more positive estimate and *z* value) in the HA group compared to the LA group across two vowel contrasts, with a stronger effect of musical aptitude in the /ɛ/−/æ/ contrast; (b) a significant interaction between group and contrast on the quadratic time term, suggesting that the FP difference follows a less U-shaped trajectory (i.e., a more negative estimate and *z* value) in the HA group than in the LA group across two vowel contrasts, with the effect of musical aptitude being more pronounced in the /ɛ/−/æ/ contrast; and (c) a significant interaction between group and contrast on the cubic time term, indicating that the FP difference exhibits a sharper S-shaped trajectory (i.e., a more negative estimate and *z* value) in the HA group compared to the LA group across two vowel contrasts, with a stronger effect of musical aptitude in the /ɛ/−/æ/ contrast. A more ascending and/or less U-shaped FP difference curve suggests faster target word recognition, characterized by increased activation of the target word and reduced activation of the competitor. In contrast, a less ascending and/or more U-shaped FP difference curve suggests slower target word recognition, driven by reduced target activation and increased competitor activation. A sharper S-shaped curve reflects fixations reaching an asymptote toward the end of the trial, likely

Figure 1. The identification curves for each group and contrast.

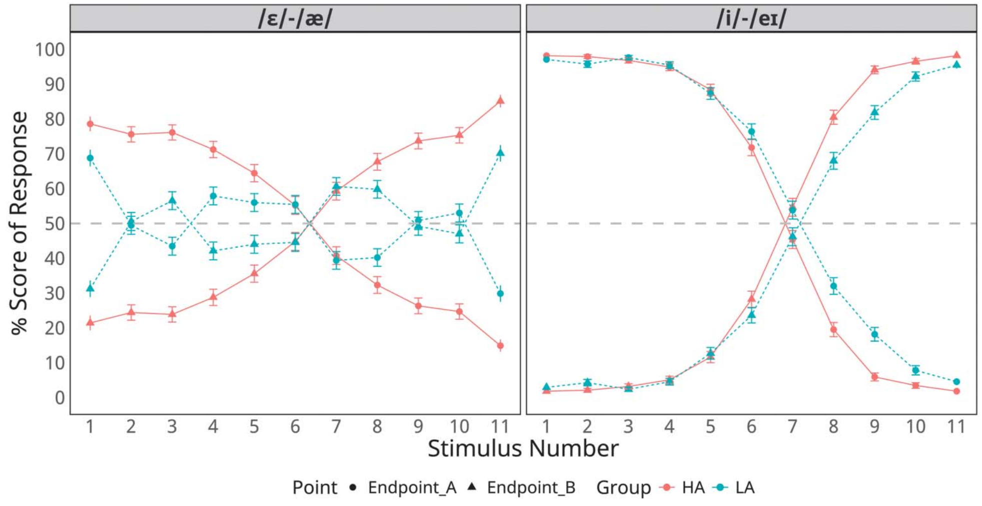
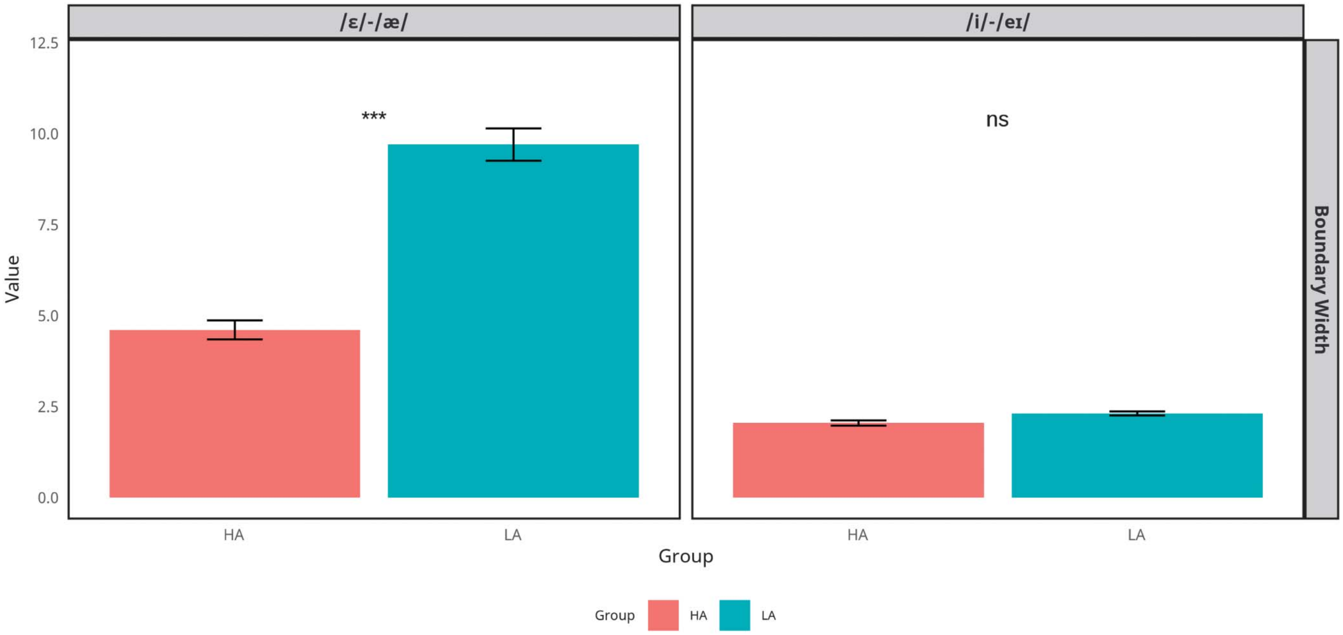


Figure 2. Boundary width for each group and contrast. Error bars represent ±1 *SD*, capturing the variability across participants. HA = high aptitude; LA = low aptitude.



indicating rapid word recognition, after which listeners shift their gaze away from the target (Mirman, 2017; Qin et al., 2019).

Figure 3 illustrates the fixation proportions across each group, contrast, and areas of interests. Figure 4 shows the FP difference within groups across the vowel contrasts of /ɛ/–/æ/ and /i/–/eɪ/. The best-fit GCA model was specified as follows: lmer(FP difference ~ (ot1 + ot2 + ot3) × Group × Contrasts + (ot1 + ot2 + ot3 | Participant) + (ot1 + ot2 + ot3 | Item), REML = FALSE). The results revealed a significant interaction between group and contrast on the linear term, $\chi^2(1) = 67.20$, $p < .001$; quadratic term, $\chi^2(1) = 134.08$, $p < .001$; and cubic term, $\chi^2(1) = 79.58$, $p < .001$. Further post hoc comparisons were conducted on the three terms, separately (see Table 2 for detailed results). It should be noted that a positive estimate for the linear term indicates a greater FP difference,

Figure 3. Fixation proportions for each group, contrast, and areas of interests (AOIs). HA = high aptitude; LA = low aptitude.

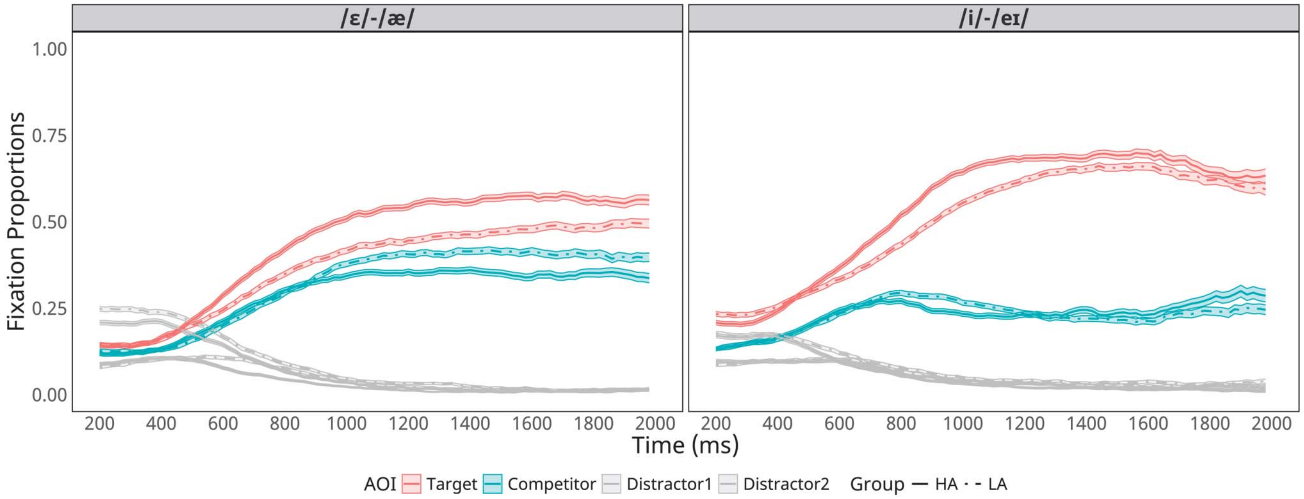
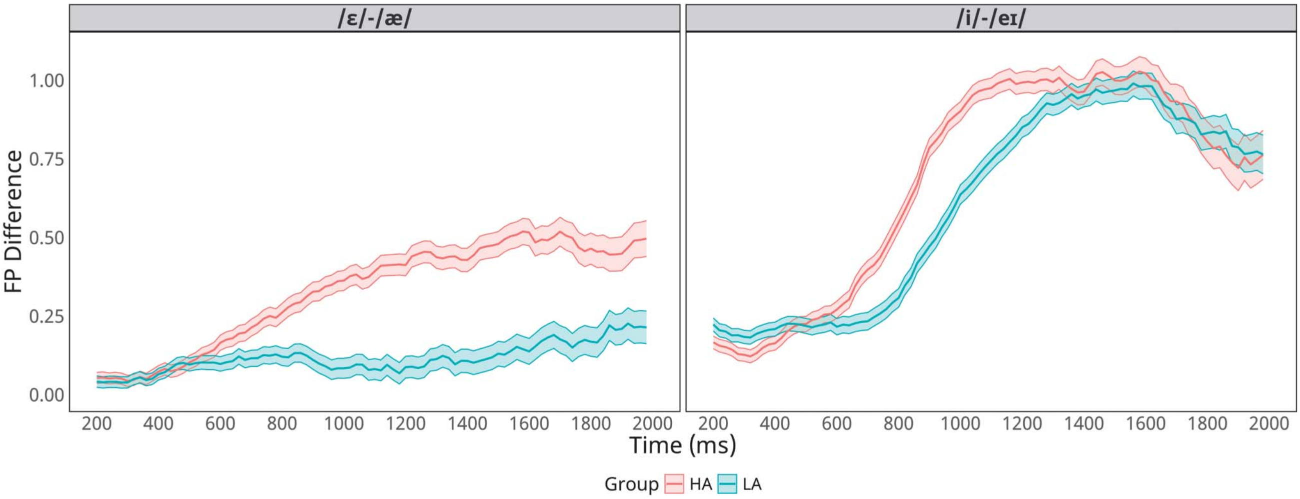


Figure 4. Fixation proportions (FP) difference for each group and contrast. HA = high aptitude; LA = low aptitude.



suggesting faster identification of the target stimuli relative to the competitor stimuli. In contrast, a positive estimate for the quadratic and cubic terms represents a smaller FP difference, reflecting slower processing and a greater competition effect (Qin et al., 2019). For comparisons between the two groups, results indicated that the HA group demonstrated significantly faster processing of the /ɛ/−/æ/ contrast compared to the LA group on both the linear term (β = 1.13, *SE* = 0.27, *z* = 4.24, *p* < .0001) and the cubic term (β = −0.71, *SE* = 0.23, *z* = −3.13, *p* = .002). Likewise, for perception of the /i/−/eɪ/ contrast, the HA group exhibited significantly faster speech processing than the LA group on the quadratic term (β = −1.59, *SE* = 0.34, *z* = −4.63, *p* < .0001). These results suggested that the HA group demonstrated more refined categorization of both vowel contrasts relative to their LA counterparts. In addition, both groups showed significant differences in perceiving the two vowel contrasts across the linear term (HA: β = −1.15, *SE* = 0.30, *z* = −3.88, *p* = .0001; LA: β = −1.99, *SE* = 0.30, *z* = −6.76, *p* < .0001) and the cubic term (HA: β = 1.15, *SE* = 0.11, *z* = 10.63, *p* < .0001; LA: β = 2.03, *SE* = 0.11, *z* = 19.29, *p* < .0001). These results indicated greater challenge in processing the /ɛ/−/æ/ contrast relative to the /i/−/eɪ/ contrast for both groups.

Correlational Results

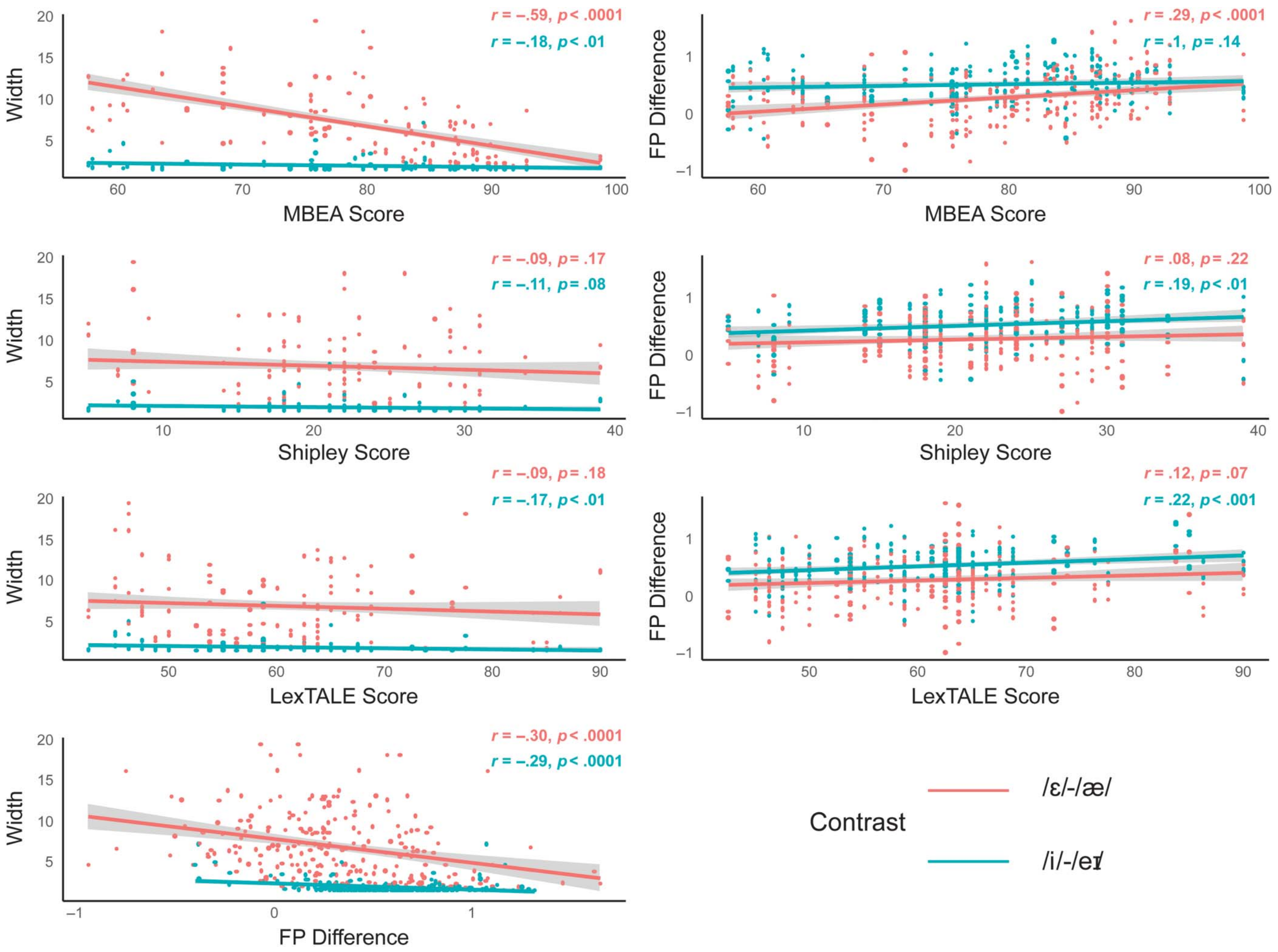
Figure 5 represents the correlational results. Specifically, the mean boundary width was significantly correlated with the mean FP difference for both vowel contrasts of /ɛ/−/æ/ (*r* = −.30, *p* < .0001) and /i/−/eɪ/ (*r* = −.29, *p* < .0001). For the /ɛ/−/æ/ contrast, the MBEA scores were strongly correlated with participants' boundary width (*r* = −.59, *p* < .0001), suggesting that participants with higher musical aptitude exhibited narrower boundary. Additionally, MBEA scores were mildly correlated with FP difference (*r* = .29, *p* < .0001), indicating that participants with higher musical aptitude showed greater differences between target and competitor fixations. In terms of the /i/−/eɪ/ contrast, boundary width was weakly correlated with both MBEA scores (*r* = −.18, *p* < .01) and LexTALE scores (*r* = −.17, *p* < .01).

Table 2. Summary of the growth curve analysis post hoc results for fixation proportions difference measurements.

		ot1				ot2				ot3			
Condition	Comparison	β	*SE*	*z*	*p*	β	*SE*	*z*	*p*	β	*SE*	*z*	*p*
/ɛ/−/æ/	HA–LA	1.13	0.27	4.24	< .0001	−0.49	0.34	−1.44	.15	−0.71	0.23	−3.13	< .01
/i/−/eɪ/	HA–LA	0.28	0.27	1.06	.29	−1.59	0.34	−4.63	< .0001	0.18	0.23	0.78	.43
HA	(/ɛ/−/æ/)-(/i/−/eɪ/)	−1.15	0.30	−3.88	< .001	0.93	0.35	2.67	.01	1.15	0.11	10.63	< .0001
LA	(/ɛ/−/æ/)-(/i/−/eɪ/)	−1.99	0.30	−6.76	< .0001	−0.17	0.35	−0.47	.64	2.03	0.11	19.29	< .0001

Note. ot1 = linear term; ot2 = quadratic term; ot3 = cubic term; HA = high aptitude; LA = low aptitude.

Figure 5. Correlation between test score and perceptual measurements for each contrast. MBEA = Montreal Battery of Evaluation of Amusia; LexTALE = Lexical Test for Advanced Learners of English; FP = fixation proportions.



Additionally, FP difference was weakly correlated with both Shipley scores (*r* = .19, *p* < .01) and LexTALE scores (*r* = .22, *p* < .001). Given that L2 English test scores were not significantly correlated with the perceptual measurements for the /ɛ/−/æ/ contrast, these results suggested that L2 English proficiency might potentially improve perception only for the /i/−/eɪ/ contrast. Overall, these results indicated that musical aptitude plays a pivotal role in categorizing the acoustically less salient /ɛ/−/æ/ contrast, which demands refined auditory sensitivity for categorization, whereas L2 English proficiency has a relatively trivial role, enhancing perception only for the /i/−/eɪ/ contrast.

Discussion

The present study employed an eye-tracking technique with VWP to examine the effect of musical aptitude on the CP of English vowels among Mandarin-speaking participants. Acoustic salience of vowel contrasts was manipulated to explore its modulations on this cross-domain transfer. The results were largely consistent with our predictions. For the effect of musical aptitude, participants with HA demonstrated more refined categorization of L2 vowel contrasts relative to their LA counterparts. For the interaction between musical aptitude and acoustic salience, a more robust transfer effect of musical aptitude was observed in the perception of the less salient /ɛ/−/æ/ contrast. To the best of our knowledge, this study represents an initial effort to systematically investigate how inherent musical aptitude influences the perception of non-native speech sounds, using Mandarin speakers' perception of English vowels as a test case. By integrating the online procedure of VWP and fine-grained measurement of CP paradigm, this study contributes to a broader understanding of cross-domain transfer from music to language.

Musical Aptitude of L2 Listeners Affecting Perception of English Vowels

The behavioral data revealed significant differences between the HA group and LA group in categorizing the /ɛ/−/æ/ contrast, with HA participants showing significantly narrower boundary width than their LA counterparts. However, this between-groups discrepancy was insignificant for identification performances of the /i/−/eɪ/ continuum, potentially attributable to the contrast's higher acoustic salience that facilitated uniformly better behavioral responses across both groups. Crucially, eye-tracking data provided a more nuanced measure on real-time perceptual processing, which confirmed significant group difference for both vowel contrasts. Specifically, the HA group maintained a greater FP difference over time compared to the LA group even for the acoustically more salient /i/−/eɪ/ contrast, exhibiting superior perceptual ability for HA listeners. These findings underscore the temporal resolution advantage of eye-tracking methodologies. In other words, by capturing temporal dynamics of participants' visual attention and incremental decision-making processes, eye-tracking can identify subtle perceptual differences between groups that could escape traditional behavioral measures on response time or accuracy. Collectively, these results indicated more refined categorization of English vowels for Mandarin-speaking participants with HA relative to their counterparts with LA, suggesting the critical role of musical aptitude in enhancing the perception of L2 vowels, which aligned with previous reports (Kempe et al., 2015; Mankel & Bidelman, 2018; C. Zhang et al., 2017). Our findings are consistent with previous research linking music to enhanced vowel perception (Bidelman & Alain, 2015; Bidelman & Krishnan, 2010; Bidelman et al., 2014; Butera, 2015; Choi et al., 2024; Cooper et al., 2016; Elmer et al., 2017; Hutka et al., 2015; Kühnis et al., 2013; Marie et al., 2011; Z. Zhang et al., 2023). Importantly, our study extends these previous findings by demonstrating that even nonmusicians with varying musical aptitude can exhibit cross-domain transfer from music to vowel perception. This suggests that musical aptitude can differentiate musically untrained individuals in vowel perception abilities.

Several theoretical proposals could contribute to tentative explanations on the transfer effect from music to vowel perception. The strong connection between music and vowel perception has been widely recognized in the literature. Both domains rely heavily on the accurate processing of spectral information—pitch in music and formant structure in vowels. Neurophysiological and imaging studies have shown that individuals with musical training exhibit heightened sensitivity to such spectral cues, often reflected in enhanced mismatch negativity responses within auditory cortical regions. Notably, these spectral processes are predominantly supported by right-hemisphere auditory structures, particularly the right supratemporal plane and planum temporale, which are critical for spectral resolution (Jäncke et al., 2002; Kühnis et al., 2013; Poeppel, 2003; Studdert-Kennedy & Shankweiler, 1981). This neural overlap provides a compelling rationale for linking music to improved vowel encoding. Much of the existing research has focused on training-induced enhancements in vowel perception among musicians (Bidelman & Alain, 2015; Bidelman & Krishnan, 2010; Bidelman et al., 2014; Butera, 2015; Elmer et al., 2017; Hutka et al., 2015; Kühnis et al., 2013; Marie et al., 2011; Z. Zhang et al., 2023). One widely cited explanation for this musician advantage is the OPERA hypothesis (Overlap, Precision, Emotion, Repetition, and Attention; Patel, 2014), which posits that musical training imposes greater demands than speech on shared neural resources. These demands—when combined with emotional engagement, extensive repetition, and focused attention—drive experience-dependent plasticity in the auditory system, ultimately resulting in enhanced vowel processing.

In contrast to studies focusing on musicians, our findings suggest that music-related advantages may also be observed in nonmusicians, depending on their musical aptitude. Specifically, Mandarin-speaking nonmusicians with higher musical aptitude performed significantly better in L2 English vowel categorization than those with lower aptitude. These results raise the possibility that preexisting musical aptitude, rather than training alone, play a key role in the link between music and enhanced vowel perception. This perspective aligns with the gene–environment interaction hypothesis proposed by Schellenberg (2015), which suggests that musical training may amplify innate differences rather than create them. Following this perspective, individual variation in musical aptitude—and related traits such as cognitive abilities and personality—may be genetically influenced and, in turn, shape one's likelihood of engaging in musical training. These preexisting traits may also underlie heightened sensitivity to speech cues, even in the absence of formal musical training. Thus, musical aptitude itself might serve as a foundational factor for the cross-domain transfer to speech processing. The current findings offer preliminary support for this account, demonstrating that vowel perception performance varied systematically with musical aptitude among nonmusicians. This is consistent with previous studies (Kempe et al., 2015; Mankel & Bidelman, 2018; C. Zhang et al., 2017), which have similarly emphasized the role of innate musical aptitude in enhancing vowel processing outcomes. Nevertheless, caution is warranted in interpreting these results: Our study did not include a direct comparison between musicians and nonmusicians across different levels of aptitude. Future research is needed to disentangle the relative contributions of musical training and inherent aptitude, particularly across varying types of phonemic contrasts.

Acoustic Salience of Vowel Contrasts Modulating the Cross-Domain Transfer

Correlational analyses on both behavioral and eye-tracking data revealed that musical aptitude played a pivotal role in categorizing the acoustically less salient /ɛ/–/æ/ contrast, rather than in categorizing the more salient counterpart of /i/–/eɪ/ contrast. These findings were consistent with our prediction that the effect of musical aptitude was mainly elicited by the acoustically less salient /ɛ/–/æ/ contrast, suggesting that the cross-domain transfer could be modulated by the acoustic salience of vowel contrasts. The interaction between musical aptitude and acoustic salience echoed previous studies showing that the cross-domain transfer from musical training to vowel perception is modulated by acoustic salience, and that stronger advantage of musical training is elicited by the acoustically less salient contrasts (Choi et al., 2024; Ong et al., 2020; Toh et al., 2023).

The interaction between musical aptitude and acoustic salience could be attributed to several potential possibilities. One explanation pertains to an auditory acuity perspective. For instance, Ong et al. (2020) found that the musician advantage emerged primarily for less salient, merging tone pairs. They argued that musicians, through extensive training, develop heightened sensitivity to fine-grained pitch differences. In contrasts with greater acoustic salience, however, this advantage may diminish as non-musicians are able to compensate—potentially by adopting alternative strategies—resulting in comparable performance levels (Choi et al., 2024). Similarly, processing acoustically less salient contrasts, such as /ɛ/–/æ/ in this study, likely required more refined auditory discrimination to categorize phonemes with subtle acoustic variations. As such, individuals with stronger auditory skills—often associated with higher musical aptitude—demonstrated superior categorization performance. This observation aligns with prior findings on the positive association between auditory perception and musical aptitude (Baldé et al., 2025; Kempe et al., 2015; Mankel & Bidelman, 2018; Schellenberg, 2015). Conversely, this pattern did not extend to acoustically more salient vowels, with HA and LA groups showing relatively more comparable performance in the categorization of /i/–/eɪ/ contrast. The robust acoustic distinctions of the two vowels might contribute to this finding, since both groups could rely on the salient cues between monophthong and diphthong to readily achieve asymptotical level in perceptual outcomes, unnecessarily resorting to refined listening sensitivity (Chen et al., 2019). It is possible that the salience of this contrast effectively compensated for the lower auditory sensitivity of the LA group (Choi et al., 2024). Further investigations could be conducted to investigate whether musical aptitude could influence processing of monophthongs with relatively more acoustic salience, such as the CP of /u/–/i/ contrast (Feng & Peng, 2023).

The other explanation could be derived from a learnability perspective. According to the phonological saliency hypothesis (Hua & Dodd, 2000) and the phonetic-magnitude hypothesis (Escudero et al., 2014), acoustic salience could determine perceptual difficulty (Qin et al., 2021) and acquisition rate of phonemes (Feng & Peng, 2023; Narayan et al., 2010). For example, Feng and Peng (2023) observed that the development of CP in Mandarin-speaking children and adolescents matured earlier and more easily for vowels and tones with greater acoustic salience than that for consonants that are less salient. In the same vein, Qin et al. (2021) found that the effect of perceptual learning of nonnative tones was asymmetrically influenced by acoustic salience, with the perception of more salient tonal contrasts obtaining greater training benefits. Therefore, the categorization of more salient contrasts could be attuned through accumulated learning experience and language exposure, which is further corroborated by the significant positive correlation between perceptual outcome of /i/–/eɪ/ contrast and L2 proficiency of the participants in this study. This interpretation is also compatible with the perceptual assimilation model (PAM) that claims nonnative phonemes are perceived on the basis of their phonological similarity to native counterparts (Best, 1994). In this study, /i/ and /eɪ/ are shared by both English and Chinese vowel inventories as distinct categories (Yang & Fox, 2017), allowing higher discriminability and better learnability for Mandarin-speaking learners. In contrast, Mandarin phonology lacks the vowel of /æ/ while labeling /ɛ/ as an allophone of its /E/ inventory (Zhou et al., 2022), resulting in perceptual assimilation of both L2 vowels into a single native category. Divergent proposal existed, however, with Sun and van Heuven (2007) positing phonemic assimilation of both English vowels into the Mandarin /a/ accompanied by systematic misidentification. The two standpoints were not necessarily incompatible, because both assumed great challenge for Mandarin speakers to acquire the /ɛ/–/æ/ distinction with the lack of a clear phonemic distinction in the native vowel inventory. This could pose the learners into a position to rely more on the refined listening skill that is closely interconnected with musical aptitude. It should be acknowledged that the vowel assimilation evidence for this study was obtained from prior literature, rather than a direct test of PAM on our participants. Future research could pay special attention to empirical research on English–Mandarin phonemic assimilations while investigating the bilinguals' speech perception. In addition, based on the LexTALE scores, the vast majority of current participants can be classified as showing English proficiency of "lower, and lower to upper intermediate," following the Common European Framework of Reference (Lemhöfer & Broersma, 2012). The relatively lower proficiency could adequately support matured categorization for the more salient contrast of L2 vowels, whereas much higher proficiency could be required for maturation of the less salient contrast of /ɛ/–/æ/, leaving a significant room for improvement for the innate musical aptitude to play a role. Future research is needed to contrast L2 learners with higher and lower proficiency to gain a better understanding of the interaction between musical aptitude and acoustic salience.

Limitations and Future Directions

Several limitations of this study should be noted that highlight the necessity for future research. First, because the present study recruited participants without any musical training experience, future investigations could consider including a group of formally trained musicians to examine whether individuals with high musical aptitude but receiving no formal training could show comparable outcomes to the musician group in vowel categorization. This additional comparison would provide valuable insights into the gene–environment interaction view (Schellenberg, 2015), disentangling the nature (i.e., musical aptitude) versus nurture (musical training) factors of music domain in affecting phonological processing in speech domain (Mankel & Bidelman, 2018). Second, this study mainly focused on the cross-domain transfer of musical aptitude to vowel in terms of perception, neglecting the production perspective. Future studies could extend this line of research to vowel production to examine whether musical aptitude could contribute to transfer effects in vowel production (Jekiel & Malarski, 2021, 2023). Such studies would add more empirical evidence for the perception–production link, shedding light on the interconnection between auditory and articulatory processes of speech. Last, the current investigation lacks direct neural evidence to elucidate the mechanisms underlying the observed far-transfer effects. Future research could integrate electrophysiological and/or neuroimaging techniques to investigate the neural correlates of the cross-domain transfer effect (Neves et al., 2022; Peretz et al., 2015; Salmi et al., 2017; Zuk et al., 2020). These neural measures could help clarify whether the transfer effects are originated from or mediated by basic auditory functions, providing deeper insights into the cognitive and neural foundations of vowel perception.

Despite these limitations, findings of this study provide valuable insights into the cross-domain transfer effects from musical aptitude to vowel perception and highlight how this transfer is modulated by acoustic salience. On the one hand, this study adds robust evidence that nonmusicians with varying levels of inherent musical aptitude can exhibit the link between music and enhanced vowel perception (Kempe et al., 2015; C. Zhang et al., 2017), showing that this link is not confined to trained musicians in previous findings. On the other hand, our findings highlight that this cross-domain transfer effect is sensitive to acoustic salience, with a stronger influence of musical aptitude observed for the acoustically less salient vowel contrast. In a nutshell, this study enhances our understanding of the cross-domain transfer effects of musical aptitude and how they are modulated in the context of vowel perception.

Conclusions

The present study utilized the eye-tracking technique with VWP to examine the influence of musical aptitude on the CP of English vowels among Mandarin-speaking adults. The results revealed that participants with higher musical aptitude significantly outperformed their lower aptitude counterparts in vowel categorization, regardless of the acoustic salience of the contrasts. Furthermore, this cross-domain transfer effect is modulated by the acoustic salience, with more robust effect of musical aptitude observed for the acoustically less salient vowel contrast. Overall, this study promoted a better understanding of cross-domain transfer effects from music to speech perception.

Data Availability Statement

The raw data supporting the conclusions of this article will be made available by the corresponding author upon request, without undue reservation.

Acknowledgments

This work was supported by grants from the Natural Science Foundation of Shandong Province (ZR2023QA117) and the Social Science Planning Project of Shandong Province (24DYYJ02), both awarded to Hao Zhang.

References

Baldé, A. M., Lima, C. F., & Schellenberg, E. G. (2025). Associations between musical expertise and auditory processing. *Journal of Experimental Psychology: Human Perception and Performance, 51*(6), 747–763. https://doi.org/10.1037/xhp0001312

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics, 29*(4), 1165–1188. https://doi.org/10.1214/aos/1013699998

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 233–277). MIT Press. https://doi.org/10.7551/mitpress/2387.003.0011

Bidelman, G. M. (2015). Induced neural beta oscillations predict categorical speech perception abilities. *Brain and Language, 141,* 62–69. https://doi.org/10.1016/j.bandl.2014.11.003

Bidelman, G. M., & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *The Journal of Neuroscience, 35*(3), 1240–1249. https://doi.org/10.1523/jneurosci.3292-14.2015

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research, 1355,* 112–125. https://doi.org/10.1016/j.brainres.2010.07.100

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience, 40*(4), 2662–2673. https://doi.org/10.1111/ejn.12627

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning, 66*(4), 774–808. https://doi.org/10.1111/lang.12159

Butera, I. M. (2015). From notes to vowels: Neural correlations between musical training and speech processing. *The Journal of Neuroscience, 35*(22), 8379–8381. https://doi.org/10.1523/jneurosci.1102-15.2015

Chen, F., & Peng, G. (2021). Categorical perception of pitch contours and voice onset time in Mandarin-speaking adolescents with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research, 64*(11), 4468–4484. https://doi.org/10.1044/2021_jslhr-20-00725

Chen, F., Zhang, H., Wang, W. S. Y., & Peng, G. (2019). Intrinsic cues and vowel categorical perception. *Linguistic Sciences, 18*(4), 410–425. https://doi.org/10.7509/j.linsci.201808.032304

Choi, W. (2022). 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*(11), 4047–4059. https://doi.org/10.1044/2022_jslhr-22-00175

Choi, W., Ling, C. L. K., & Wu, C. H. J. (2024). Musical advantage in lexical tone perception hinges on musical instrument: A comparison between pitched musicians, unpitched musicians, and nonmusicians. *Music Perception, 41*(5), 360–377. https://doi.org/10.1525/mp.2024.41.5.360

Chui, Y. T., & Qin, Z. (2024). Distributional learning and overnight consolidation of nonnative tonal contrasts by tonal language speakers. *Journal of Speech, Language, and Hearing Research, 67*(7), 2038–2052. https://doi.org/10.1044/2024_jslhr-23-00711

Connell, K., Hüls, S., Martínez-García, M. T., Qin, Z., Shin, S., Yan, H., & Tremblay, A. (2018). English learners' use of segmental and suprasegmental cues to stress in lexical access: An eye-tracking study. *Language Learning, 68*(3), 635–668. https://doi.org/10.1111/lang.12288

Cooper, A., Wang, Y., & Ashley, R. (2016). Thai rate-varied vowel length perception and the impact of musical experience. *Language and Speech, 60*(1), 65–84. https://doi.org/10.1177/0023830916642489

Creel, S. C. (2014). Tipping the scales: Auditory cue weighting changes over development. *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 1146–1160. https://doi.org/10.1037/a0036057

Cui, A., & Kuang, J. (2019). The effects of musicality and language background on cue integration in pitch perception. *The Journal of the Acoustical Society of America, 146*(6), 4086–4096. https://doi.org/10.1121/1.5134442

Davies, M. (2008) *The Corpus of Contemporary American English (COCA)* [Data set]. https://www.english-corpora.org/coca/

Dial, H. R., McMurray, B., & Martin, R. C. (2019). Lexical processing depends on sublexical processing: Evidence from the visual world paradigm and aphasia. *Attention, Perception, & Psychophysics, 81*(4), 1047–1064. https://doi.org/10.3758/s13414-019-01718-3

Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis* [R package]. http://www.eyetrackingr.com

Elmer, S., Greber, M., Pushparaj, A., Kühnis, J., & Jäncke, L. (2017). Faster native vowel discrimination learning in musicians is mediated by an optimization of mnemonic functions. *Neuropsychologia, 104,* 64–75. https://doi.org/10.1016/j.neuropsychologia.2017.08.001

Escudero, P., Best, C. T., Kitamura, C., & Mulak, K. E. (2014). Magnitude of phonetic distinction predicts success at early word learning in native and non-native accents. *Frontiers in Psychology, 5,* Article 1059. https://doi.org/10.3389/fpsyg.2014.01059

Feng, Y., & Peng, G. (2023). Development of categorical speech perception in Mandarin-speaking children and adolescents. *Child Development, 94*(1), 28–43. https://doi.org/10.1111/cdev.13837

Finney, D. J. (1971). *Probit analysis* (3rd ed.). Cambridge University Press.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5*(4), 171–189. https://doi.org/10.1177/002383096200500401

Ghaffarvand Mokari, P., & Werner, S. (2018). Perceptual training of second-language vowels: Does musical ability play a role? *Journal of Psycholinguistic Research, 47*(1), 95–112. https://doi.org/10.1007/s10936-017-9517-8

Gordon, E. E. (1989). *Advanced measures of music audiation*. Gia Publications.

Gordon, E. E. (2007). *Learning sequences in music: A contemporary music learning theory*. Gia Publications.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.11–10.112). Wiley.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 1–52). Cambridge University Press.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27*(1), 3–42. https://doi.org/10.1017/S030500099900402X

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460–482. https://doi.org/10.1016/j.jml.2007.02.001

Hui, N.-Y., Yuan, M., Fong, M. C.-M., & Wang, W. S. (2020). L2 proficiency predicts inhibitory ability in L1-dominant speakers. *International Journal of Bilingualism, 24*(5–6), 984–998. https://doi.org/10.1177/1367006920914399

Hutka, S., Bidelman, G. M., & Moreno, S. (2015). Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. *Neuropsychologia, 71,* 52–63. https://doi.org/10.1016/j.neuropsychologia.2015.03.019

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods, 55*(7), 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98,* 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Jansen, N., Harding, E. E., Loerts, H., Başkent, D., & Lowie, W. (2023). The relation between musical abilities and speech prosody perception: A meta-analysis. *Journal of Phonetics, 101,* Article 101278. https://doi.org/10.1016/j.wocn.2023.101278

Jäncke, L., Wüstenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *NeuroImage, 15*(4), 733–746. https://doi.org/10.1006/nimg.2001.1027

Jekiel, M., & Malarski, K. (2021). Musical hearing and musical experience in second language English vowel acquisition. *Journal of Speech, Language, and Hearing Research, 64*(5), 1666–1682. https://doi.org/10.1044/2021_jslhr-19-00253

Jekiel, M., & Malarski, K. (2023). Musical hearing and the acquisition of foreign-language intonation. *Studies in Second Language Learning and Teaching, 13*(1), 151–178. https://doi.org/10.14746/ssllt.23166

Kawahara, H., & Morise, M. (2011). Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework. *Sadhana, 36*(5), 713–727. https://doi.org/10.1007/s12046-011-0043-3

Kempe, V., Bublitz, D., & Brooks, P. J. (2015). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology, 106*(2), 349–366. https://doi.org/10.1111/bjop.12092

Kragness, H. E., Swaminathan, S., Cirelli, L. K., & Schellenberg, E. G. (2021). Individual differences in musical ability are stable over time in childhood. *Developmental Science, 24*(4), Article e13081. https://doi.org/10.1111/desc.13081

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Kühnis, J., Elmer, S., Meyer, M., & Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia, 51*(8), 1608–1618. https://doi.org/10.1016/j.neuropsychologia.2013.04.007

Ladefoged, P., & Johnson, K. (2006). *A course in phonetics* (5th ed.). Thomson Wadsworth.

Law, L. N., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills. *PLOS ONE, 7*(12), Article e52508. https://doi.org/10.1371/journal.pone.0052508

Lehmann, A., Skoe, E., Moreau, P., Peretz, I., & Kraus, N. (2015). Impairments in musical abilities reflected in the auditory brainstem: Evidence from congenital amusia. *European Journal of Neuroscience, 42*(1), 1644–1650. https://doi.org/10.1111/ejn.12931

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *emmeans: Estimated marginal means, aka least-squares means* (R Package Version 4.0-3). https://cran.r-project.org/package=emmeans

Lewis, G. A., & Bidelman, G. M. (2020). Autonomic nervous system correlates of speech categorization revealed through pupillometry. *Frontiers in Neuroscience, 13,* Article 1418. https://doi.org/10.3389/fnins.2019.01418

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368. https://doi.org/10.1037/h0044417

Loui, P., Li, H. C., Hohmann, A., & Schlaug, G. (2011). Enhanced cortical connectivity in absolute pitch musicians: A model for local hyperconnectivity. *Journal of Cognitive Neuroscience, 23*(4), 1015–1026. https://doi.org/10.1162/jocn.2010.21500

Ma, J., Zhu, J., Yao, X., & Chen, Y. (2024). Categorical perception of lexical tones and stops in Mandarin-speaking musicians and nonmusicians. *SAGE Open, 14*(1). https://doi.org/10.1177/21582440241227703

Mankel, K., Barber, J., & Bidelman, G. M. (2020). Auditory categorical processing for speech is modulated by inherent musical listening skills. *Neuroreport, 31*(2), 162–166. https://doi.org/10.1097/wnr.0000000000001369

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences, 115*(51), 13129–13134. https://doi.org/10.1073/pnas.1811793115

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience, 23*(10), 2701–2715. https://doi.org/10.1162/jocn.2010.21585

McHaney, J. R., Tessmer, R., Roark, C. L., & Chandrasekaran, B. (2021). Working memory relates to individual differences in speech category learning: Insights from computational modeling and pupillometry. *Brain and Language, 222,* Article 105010. https://doi.org/10.1016/j.bandl.2021.105010

McMurray, B. (2023). I'm not sure that curve means what you think it means: Toward a [more] realistic understanding of the role of eye-movement generation in the visual world paradigm. *Psychonomic Bulletin & Review, 30*(1), 102–146. https://doi.org/10.3758/s13423-022-02143-8

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance, 34*(6), 1609–1631. https://doi.org/10.1037/a0011747

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology, 54*(8), 1472–1491. https://doi.org/10.1037/dev0000542

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–B42. https://doi.org/10.1016/s0010-0277(02)00157-9

Mirman, D. (2017). *Growth curve analysis and visualization using R*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315373218

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Musso, M., Fürniss, H., Glauche, V., Urbach, H., Weiller, C., & Rijntjes, M. (2020). Musicians use speech-specific areas when processing tones: The key to their superior linguistic competence? *Behavioural Brain Research, 390,* Article 112662. https://doi.org/10.1016/j.bbr.2020.112662

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science, 13*(3), 407–420. https://doi.org/10.1111/j.1467-7687.2009.00898.x

Neves, L., Correia, A. I., Castro, S. L., Martins, D., & Lima, C. F. (2022). Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. *Neuroscience & Biobehavioral Reviews, 140,* Article 104777. https://doi.org/10.1016/j.neubiorev.2022.104777

Oikkonen, J., Huang, Y., Onkamo, P., Ukkola-Vuoti, L., Raijas, P., Karma, K., Vieland, V. J., & Järvelä, I. (2015). A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. *Molecular Psychiatry, 20*(2), 275–282. https://doi.org/10.1038/mp.2014.8

Ong, J. H., Wong, P. C. M., & Liu, F. (2020). Musicians show enhanced perception, but not production, of native lexical tones. *The Journal of the Acoustical Society of America, 148*(6), 3443–3454. https://doi.org/10.1121/10.0002776

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research, 308,* 98–108. https://doi.org/10.1016/j.heares.2013.08.011

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics, 38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., Paquette, S., & Bouchard, B. (2008). On-line identification of congenital amusia. *Music Perception, 25*(4), 331–343. https://doi.org/10.1525/mp.2008.25.4.331

Peretz, I., Vuvan, D., Lagrois, M. É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1664), Article 20140090. https://doi.org/10.1098/rstb.2014.0090

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 23*(1), Article 148. https://doi.org/10.1121/1.1917300

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition, 3*(1), 7–18. https://doi.org/10.3758/bf03198202

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time.' *Speech Communication, 41*(1), 245–255. https://doi.org/10.1016/s0167-6393(02)00107-3

Qin, Z., Jin, R., & Zhang, C. (2022). The effects of training variability and pitch aptitude on the overnight consolidation of lexical tones. *Journal of Speech, Language, and Hearing Research, 65*(9), 3377–3391. https://doi.org/10.1044/2022_jslhr-22-00058

Qin, Z., Tremblay, A., & Zhang, J. (2019). Influence of within-category tonal information in the recognition of Mandarin-Chinese words by native and non-native listeners: An eye-tracking study. *Journal of Phonetics, 73,* 144–157. https://doi.org/10.1016/j.wocn.2019.01.002

Qin, Z., Zhang, C., & Wang, W. S. (2021). The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *The Journal of the Acoustical Society of America, 149*(1), 435–446. https://doi.org/10.1121/10.0003330

Qin, Z., & Zhang, J. (2024). The role of coarticulatory tonal information in Cantonese spoken word recognition: An eye-tracking study. *Linguistics Vanguard, 10*(1), 81–91. https://doi.org/10.1515/lingvan-2022-0158

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance, 5*(1), 129–145. https://doi.org/10.1037/0096-1523.5.1.129

Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica, 138*(1), 1–10. https://doi.org/10.1016/j.actpsy.2011.03.007

Salmi, J., Koistinen, O.-P., Glerean, E., Jylänki, P., Vehtari, A., Jääskeläinen, I. P., Mäkelä, S., Nummenmaa, L., Nummi-Kuisma, K., Nummi, I., & Sams, M. (2017). Distributed neural signatures of natural audiovisual speech and music in the human auditory cortex. *NeuroImage, 157,* 108–117. https://doi.org/10.1016/j.neuroimage.2016.12.005

Schellenberg, E. G. (2015). Music training and speech perception: A gene–environment interaction. *Annals of the New York Academy of Sciences, 1337*(1), 170–177. https://doi.org/10.1111/nyas.12627

Schellenberg, E. G., & Lima, C. F. (2024). Music training and nonmusical abilities. *Annual Review of Psychology, 75*(1), 87–128. https://doi.org/10.1146/annurev-psych-032323-051354

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience, 5*(7), 688–694. https://doi.org/10.1038/nn871

Shahin, A. J., Roberts, L. E., Pantev, C., Aziz, M., & Picton, T. W. (2007). Enhanced anterior-temporal processing for complex tones in musicians. *Clinical Neurophysiology, 118*(1), 209–220. https://doi.org/10.1016/j.clinph.2006.09.019

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Journal of Psychology, 9*(2), 371–377. https://doi.org/10.1080/00223980.1940.9917704

Smit, E. A., Milne, A. J., & Escudero, P. (2022). Music perception abilities and ambiguous word learning: Is there cross-domain transfer in nonmusicians? *Frontiers in Psychology, 13,* Article 801263. https://doi.org/10.3389/fpsyg.2022.801263

Stevens, K. N., Libermann, A. M., Studdert-Kennedy, M., & Öhman, S. E. G. (1969). Crosslanguage study of vowel perception. *Language and Speech, 12*(1), 1–23. https://doi.org/10.1177/002383096901200101

Strait, D. L., Hornickel, J., & Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral and Brain Functions, 7*(1), Article 44. https://doi.org/10.1186/1744-9081-7-44

Studdert-Kennedy, M., & Shankweiler, D. (1981). Hemispheric specialization for language processes. *Science, 211*(4485), 960–961. https://doi.org/10.1126/science.7466372

Sun, L., & van Heuven, V. J. (2007). Perceptual assimilation of English vowels by Chinese listeners: Can native-language interference be predicted? *Linguistics in the Netherlands, 24*(1), 150–161. https://doi.org/10.1075/avt.24.15sun

Swaminathan, S., & Schellenberg, E. G. (2017). Musical competence and phoneme perception in a foreign language. *Psychonomic Bulletin & Review, 24*(6), 1929–1934. https://doi.org/10.3758/s13423-017-1244-5

Toh, X. R., Tan, S. H., Wong, G., Lau, F., & Wong, F. C. K. (2023). Enduring musician advantage among former musicians in prosodic pitch perception. *Scientific Reports, 13*(1), Article 2657. https://doi.org/10.1038/s41598-023-29733-3

Turner, J. (2022). Analysing the relationship between L2 production and different stages of L2 processing: Eye-tracking and acoustic evidence for a novel contrast. *Journal of Phonetics, 91,* Article 101134. https://doi.org/10.1016/j.wocn.2022.101134

Ullén, F., Mosing, M. A., Holm, L., Eriksson, H., & Madison, G. (2014). Psychometric properties and heritability of a new online test for musicality, the Swedish musical discrimination test. *Personality and Individual Differences, 63,* 87–93. https://doi.org/10.1016/j.paid.2014.01.057

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences, 20*(3), 188–196. https://doi.org/10.1016/j.lindif.2010.02.004

Yang, J., & Fox, R. A. (2017). L1–L2 interactions of vowel systems in young bilingual Mandarin–English children. *Journal of Phonetics, 65,* 60–76. https://doi.org/10.1016/j.wocn.2017.06.002

Yao, Y., Chen, X., Chen, F., & Zhu, J. (2022). Musical training enhances categorical perception of speech in preschoolers: Training duration and musical program matter. *Journal of Speech, Language, and Hearing Research, 65*(11), 4469–4484. https://doi.org/10.1044/2022_jslhr-22-00216

Yashaswini, L., & Maruthy, S. (2020). Effect of music training on categorical perception of speech and music. *Journal of Audiology & Otology, 24*(3),140–148. https://doi.org/10.7874/jao.2019.00500

Zhang, C., Shao, J., & Huang, X. (2017). Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. *PLOS ONE, 12*(8), Article e0183151. https://doi.org/10.1371/journal.pone.0183151

Zhang, H., Chen, F., Yan, N., Wang, L., Shi, F., & Ng, M. L. (2016). The influence of language experience on the categorical perception of vowels: Evidence from Mandarin and Korean. In *Proceedings of Interspeech 2016* (pp. 873–877). https://doi.org/10.21437/Interspeech.2016-887

Zhang, H., Dai, X., Ma, W., Ding, H., & Zhang, Y. (2024). Investigating perception to production transfer in children with cochlear implants: A high variability phonetic training study. *Journal of Speech, Language, and Hearing Research, 67*(4), 1206–1228. https://doi.org/10.1044/2023_jslhr-23-00573

Zhang, H., Ma, W., Ding, H., & Zhang, Y. (2023). Sustainable benefits of high variability phonetic training in Mandarin-speaking kindergarteners with cochlear implants: Evidence from categorical perception of lexical tones. *Ear and Hearing, 44*(5), 990–1006. https://doi.org/10.1097/aud.0000000000001341

Zhang, K., Tao, R., & Peng, G. (2023). The advantage of the music-enabled brain in accommodating lexical tone variabilities. *Brain and Language, 247,* Article 105348. https://doi.org/10.1016/j.bandl.2023.105348

Zhang, Y. (2016). Categorical perception. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, & J. Myers (Eds.), *Encyclopedia of Chinese language and linguistics* (Vol. 1, pp. 126–130). Brill. https://doi.org/10.1163/2210-7363_ecll_COM_000071

Zhang, Z., Zhang, H., Sommer, W., Yang, X., Wei, Z., & Li, W. (2023). Musical training alters neural processing of tones and vowels in classic Chinese poems. *Brain and Cognition, 166,* Article 105952. https://doi.org/10.1016/j.bandc.2023.105952

Zheng, C., Saito, K., & Tierney, A. (2022). Successful second language pronunciation learning is linked to domain-general auditory processing rather than music aptitude. *Second Language Research, 38*(3), 477–497. https://doi.org/10.1177/0267658320978493

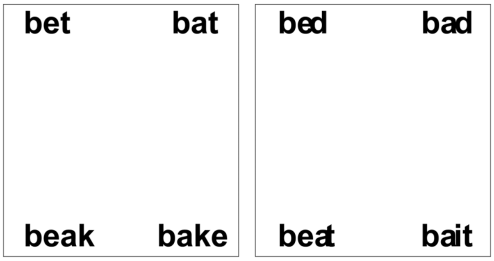
Zhou, A., Dmitrieva, O., & Olson, D. J. (2022). The effect of allophonic variability on L2 contrast perception: Evidence from perception of English vowels. *JASA Express Letters, 2*(12), Article 125201. https://doi.org/10.1121/10.0016602

Zhu, J., Chen, X., & Yang, Y. (2021). Effects of amateur musical experience on categorical perception of lexical tones by native Chinese adults: An ERP study. *Frontiers in Psychology, 12,* Article 611189. https://doi.org/10.3389/fpsyg.2021.611189

Zuk, N. J., Teoh, E. S., & Lalor, E. C. (2020). EEG-based classification of natural sounds reveals specialized responses to speech and music. *NeuroImage, 210,* Article 116558. https://doi.org/10.1016/j.neuroimage.2020.116558

Appendix

Visual Displays of the Eye-Tracking Trials in Blocks 1 and 2 Following the Visual World Paradigm


bet
bat
beak
bake
bed
bad
beat
bait