

MUSICAL ADVANTAGE IN LEXICAL TONE PERCEPTION HINGES ON MUSICAL INSTRUMENT: A COMPARISON BETWEEN PITCHED MUSICIANS, UNPITCHED MUSICIANS, AND NONMUSICIANS

WILLIAM CHOI, CHEUK LAM KATIE LING, &
CHUN HIM JASON WU
The University of Hong Kong, Pokfulam, Hong Kong

DIFFERENT MUSICAL INSTRUMENTS HAVE DIFFERENT pitch processing demands. However, correlational studies have seldom considered the role of musical instruments in music-to-language transfer. Addressing this research gap could contribute to a nuanced understanding of music-to-language transfer. To this end, we investigated whether pitched musicians had a unique musical advantage in lexical tone perception relative to unpitched musicians and nonmusicians. Specifically, we compared Cantonese pitched musicians, unpitched musicians, and nonmusicians on Thai tone discrimination and sequence recall. In the Thai tone discrimination task, the pitched musicians outperformed the unpitched musicians and the nonmusicians. Moreover, the unpitched musicians and the nonmusicians performed similarly. In the Thai tone sequence recall task, both pitched and unpitched musicians recalled level tone sequences more accurately than the nonmusicians, but the pitched musicians showed the largest musical advantage. However, the three groups recalled contour tone sequences with similar accuracy. Collectively, the pitched musicians had a unique musical advantage in lexical tone discrimination and the largest musical advantage in level tone sequence recall. From a theoretical perspective, this study offers correlational evidence for the Precision element of the OPERA hypothesis. The choice of musical instrumental may matter for music-to-language transfer in lexical tone discrimination and level tone sequence recall.

Received: June 28, 2023, accepted January 29, 2024.

Key words: pitch, perception, speech, tonality, music and language

MUSIC TRAINING ENHANCES PHONOLOGICAL awareness (Gordon et al., 2015; Tierney & Kraus, 2014), speech-in-noise perception

(Coffey et al., 2017; Hennessy et al., 2022; Maillard et al., 2023), and lexical tone perception (Nan et al., 2018; Patel, 2014). In correlational research, musicians often perceive lexical tones more accurately than nonmusicians (Alexander et al., 2005; Choi, 2020; Kraus & Chandrasekaran, 2010), reflecting an acquired or pre-existing advantage in lexical tone perception (Patel, 2014; Schellenberg, 2015). Unlike many music perception studies (e.g., Shahin et al., 2003; Slater & Kraus, 2015; Tervaniemi et al., 2016), most lexical tone perception studies have not considered the heterogeneity of musicianship (e.g., Choi, 2020; Lee et al., 2014; Zheng & Samuel, 2018). Specifically, they have largely represented musicianship as a binary variable (i.e., musician or nonmusician). Different musical instruments have different pitch processing demands, so it is possible that not all types of musicians exhibit an advantage in lexical tone perception. To extend the previous lexical tone perception studies, we compared the lexical tone perception abilities of pitched musicians (i.e., violinists and pianists), unpitched musicians (i.e., unpitched percussionists), and nonmusicians.

According to the OPERA hypothesis, long-term music training would facilitate speech perception upon the fulfillment of five conditions—**Overlap**, **Precision**, **Emotion**, **Repetition**, and **Attention** (Patel, 2011, 2014). Music and language perceptual attributes (e.g., musical pitch and lexical tones) often share a common acoustic feature (e.g., periodicity). For **Overlap**, although listeners may process music and language perceptual attributes differently at the cortical level, they recruit overlapping subcortical networks to process their common acoustic feature. For **Precision**, music training must place a higher processing demand on the acoustic feature than does speech. For **Emotion**, the music training must bring about strong positive emotions. For **Repetition**, the music training must be repeated frequently. For **Attention**, the music training must require focused attention. When these conditions are met, long-term musical experience will enhance the subcortical processing of acoustic feature (e.g., periodicity) shared by the music and language perceptual attributes (e.g., musical pitch and lexical tones). Then, such enhancement

would feed forward to benefit the perception of the language perceptual attribute (e.g., lexical tones).

Consistent with OPERA, there is causal evidence of music-to-language transfer in lexical tone discrimination. In a randomized controlled trial (RCT) study, children were assigned to music training or painting training for six months (Moreno et al., 2009). Compared to their pre-training baseline, only the music training group improved on behavioral and neural pitch deviance detection in speech. In an RCT study that directly assessed lexical tone perception, children were assigned to music training, reading training, or a no-contact control group for six months (Nan et al., 2018). Relative to their pre-training baseline, only the music training group showed enhanced positive mismatch responses (pMMRs) to lexical tone violations. This reflected the positive effect of music training on children's neuronal sensitivity to lexical tones.

Supplementing the causal evidence, correlational studies showed that musicians outperformed nonmusicians in lexical tone discrimination (e.g., Burnham et al., 2015; Choi, 2020; Delogu et al., 2010). Prior to our review, it is important to note that the correlational research design has limited causal inference. On the one hand, it is possible that the musicians outperform the nonmusicians because they have received music training (Patel, 2011, 2014; Tierney & Kraus, 2014). On the other hand, it is also possible that music training exerts no effect or at best only serves to exaggerate the pre-existing differences between musicians and nonmusicians (Schellenberg, 2015). Despite the caveat of correlational design, it can investigate the effects of long-term musical experience (i.e., lasting more than 6 years, Zhang et al., 2020), which are typically infeasible in RCTs (Gordon et al., 2015).

Here, we summarize previous correlational findings as suggestive evidence of music-to-language transfer in tone discrimination (Burnham et al., 2015; Choi, 2020; Delogu et al., 2010), without prejudice to its alternative interpretation (Schellenberg, 2015). An early study tested Italian musicians, nonmusicians, and Mandarin nonmusicians on lexical tone and segmental discrimination (Delogu et al., 2010). Although Italian musicians and nonmusicians discriminated segmental information with similar accuracy, the former outperformed the latter on Mandarin tone discrimination. Moreover, Italian musicians even performed on a par with Mandarin nonmusicians. This has suggested that musicianship is associated with a lexical tone-specific perceptual benefit instead of a general one that applies to any speech information (e.g., vowels and consonants). A later study tested the effects of absolute pitch and musicianship on Thai tone discrimination (Burnham et al., 2015).

Consistent with music-to-language transfer, Australian-English musicians discriminated Thai tones more accurately than did their nonmusician counterparts. Furthermore, the musicians with absolute pitch ability outperformed those without. In a more recent study, English musicians discriminated Cantonese tones more accurately than English nonmusicians only in half of the tone contexts (Choi, 2020). This has reflected that certain lexical tones may have specific acoustic features that are more relevant to musical experience. Based on this, Choi concluded that musical advantage was not general but selective to certain lexical tones.

Beyond lexical tone discrimination, correlational evidence suggested possible music-to-language transfer in non-native lexical tone identification (Alexander et al., 2005; Lee et al., 2014). Unlike in the discrimination task, participants were typically given a short tutorial on the non-native lexical tones prior to the experiment (see Alexander et al., 2005). During the experiment, participants were required to match an audibly presented lexical tone with one of the several pictures that visualized the F0 profiles of the non-native lexical tones. Consistent with the lexical tone discrimination studies, English musicians outperformed English nonmusicians on identifying Mandarin tones (Alexander et al., 2005; Lee et al., 2014). Furthermore, English musicians even achieved a native-like accuracy on Mandarin tone discrimination.

Besides lexical tone discrimination and identification, musicians also exhibit an advantage in non-native lexical tone sequence recall and word learning. A previous study assessed the abilities of English musicians and nonmusicians in recalling Cantonese level tone sequences and contour tone sequences (Choi, 2020). While both groups performed similarly on recalling Cantonese level tone sequences, English musicians recalled Cantonese contour tone sequences more accurately than did nonmusicians. In a Mandarin tone word learning study, English musicians outperformed English nonmusicians on identifying Mandarin tone words after training, despite its small sample size ($n = 17$) (Wong & Perrachione, 2007). With a larger sample size ($n = 54$), Cooper and Wang (2012) found that English musicians achieved a higher accuracy on Cantonese tone word learning than English nonmusicians. Overall, Thai musicians did not significantly outperform Thai nonmusicians on Cantonese tone word learning, suggesting that musicianship might not benefit tone word learning in tone-language speakers (e.g., *tonal listeners* hereafter). Yet, both English and Thai musicians obtained higher accuracies in pre- and post-training lexical tone identification tasks than did the nonmusicians (Cooper & Wang, 2012). In sum, the above studies provided

a crucial idea that musicians exhibit an advantage in non-native lexical tone discrimination, identification, sequence recall, and word learning, especially for non-tone-language listeners.

Collectively, previous research has suggested possible music-to-language transfer in different perceptual modes (e.g., Burnham et al., 2015; Choi, 2020; Lee et al., 2014). According to the Automatic Selective Perception model (ASP; Strange, 2011), task and stimuli nature can drive listeners to vary the balance between the phonetic and phonological modes of perception. In the phonetic mode of perception, listeners attend to concrete phonetic differences of the stimuli. In the phonological mode, listeners attend to abstract invariant phonological differences. For example, even though the same phoneme (produced by two speakers) are acoustically different, listeners still perceive them as the same phoneme. According to recent studies, listeners are more likely to engage in phonological perception when a task is abstract and involves speaker variability (Chen et al., 2019, 2023). Concerning music-to-language transfer, some previous studies only included a discrimination task with stimuli produced by a single speaker (e.g., Burnham et al., 2015). Based on ASP, listeners could simply adopt a phonetic mode (or even a psychoacoustic mode, Hallé et al., 2004) of perception and discriminate the lexical tones based on their crude F0 differences.

To better investigate music-to-language transfer in the context of phonological perception, we modified a discrimination task and adopted the sequence recall task. In a typical AXB discrimination task, listeners hear three audio stimuli produced by the same talker. They then judge whether the first (A) or third sound (B) differed from the second sound (X). In our AXB discrimination task, X and AB were produced by different talkers. Different talkers have different vocal tract size, anatomy, and motor control (Johnson et al., 1993). As such, even though our A (or B) and X were phonologically identical, they were acoustically distinct. To discriminate the lexical tones, listeners had to extract the invariant phonological information rather than the variant phonetic/acoustic information. That being said, Chen et al. (2023) argued that the discrimination task inherently orientates listeners to detect phonetic/acoustic variations. Thus, we supplemented the discrimination task with the sequence recall task.

The sequence recall task has been used extensively to assess the phonological mode of perception (e.g., Choi, 2020; Dupoux et al., 2010; Kim & Tremblay, 2021). The sequence recall task contains a familiarization phase and a testing phase. In the familiarization phase, listeners explicitly associate two sounds (e.g., /da1/ and /da2/)

with two separate keys (e.g., [1] and [2]). In the testing phase, listeners first hear a sequence of sounds (e.g., /da1-da2-da1-da2-da2/) then recall it by pressing the corresponding keys (e.g., [1][2][1][2][2]). According to Dupoux et al. (2010, p. 268), the sequence recall task assesses the ability to represent and store the non-native speech contrast in a “short-term memory phonological store.” Cognitively, the task has a higher memory load than the discrimination task. Based on previous studies, a high memory load and talker variability would drive listeners to adopt a phonological mode of perception (Chen et al., 2019, 2023). In addition to the memory load, we incorporated talker variability into our sequence recall task. In each sequence, adjacent tones were produced by different talkers. To provide response, our listeners had to engage in higher-level perceptual operations including speaker normalization, phonological encoding, and memory sequencing.

Does the Choice of Musical Instrument Matter?

Although previous studies found causal and correlational evidence of music-to-language transfer in lexical tone perception (e.g., Burnham et al., 2015; Delogu et al., 2010; Nan et al., 2018), one important component is missing: does the type of musicianship matter for music-to-language transfer? In the OPERA framework, the Precision requirement is that music training must entail more precise processing of the common acoustic feature of the music and language perceptual attributes (Patel, 2011, 2014). Accordingly, music training should enhance lexical tone perception only if it has a higher demand on pitch than does speech. As described below, different musical instruments have different processing demands on pitch. However, existing lexical tone perception studies have not considered this factor when examining music-to-language transfer (e.g., Choi, 2020; Cooper & Wang, 2012; Zheng & Samuel, 2018). Most often, the musician group contained a heterogeneous sample of musicians who had learnt musical instruments with different pitch processing demands (e.g., piano and drum).

Unlike lexical tone perception studies, many music perception studies have considered the heterogeneity of musicianship such as genre (Tervaniemi et al., 2016) and the type of musical instrument (Cameron & Grahn, 2014; Cicchini et al., 2012). Concerning genre, the timing of the last sound in a melodic sequence has a larger expressive importance in classical music and jazz relative to rock (Tervaniemi et al., 2016). Compared with rock musicians, classical and jazz musicians showed a stronger mismatch negativity (MMN) response to timing delay in the end of melodic

sequences (Tervaniemi et al., 2016). Regarding the type of musical instrument, percussion instruments (e.g., drums) have a higher demand on rhythm than non-percussion instruments (e.g., violin) (Vuust et al., 2012). As expected, percussionist musicians outperformed non-percussionist musicians on musical rhythmic perception (Cicchini et al., 2012; cf. Slater & Kraus, 2015) and reproduction (Cameron & Grahn, 2014).

Of interest in this study is pitch. Based on pitch processing demand, musical instruments can be broadly classified as *pitched* and *unpitched musical instruments*. Pitched musical instruments require players to control the pitch over an equal-tempered scale (e.g., piano) or continuously monitor and manipulate a constantly changing pitch (e.g., violin). Relative to nonmusicians, pianists and violinists showed larger auditory evoked potential to musical pitch (Shahin et al., 2003), reflecting enhanced musical pitch perception. Thus, we included pianists and violinists in the pitched musician group. Focusing on rhythm, unpitched musical instruments produce sounds of indefinite pitch and therefore have minimal pitch processing demand (e.g., bass drum and cymbal) (Alexander et al., 2005; Parker, 1983).

To extend the previous studies, we investigated whether pitched musicians exhibit a unique advantage in tone perception relative to unpitched musicians and nonmusicians. According to OPERA, enhanced subcortical pitch processing undergirds the music advantage in lexical tone perception. Since pitched musical instruments have a high pitch processing demand, we hypothesized that the pitched musicians would outperform the unpitched musicians and the nonmusicians. As unpitched musical instruments have a minimal pitch processing demand, we predicted that the unpitched musicians and nonmusicians would perform similarly. Unlike in the previous studies, our nuanced classification of pitched and unpitched musicians has enabled us to comprehensively test the Precision element of OPERA in the correlational context (e.g., Choi, 2020; Lee et al., 2014; Zheng & Samuel, 2018).

In short, we compared pitched musicians, unpitched musicians, and nonmusicians on lexical tone discrimination and sequence recall. According to meta-analyses, music training enhances cognitive skills such as working memory and nonverbal intelligence (Bigand & Tillmann, 2022; Talamini et al., 2017). Since AXB discrimination and especially sequence recall have a high cognitive demand, we statistically controlled for these factors (Chen et al., 2020). If the Precision element holds, the pitched musicians should outperform the unpitched musicians and the nonmusicians on lexical tone perception. Moreover, the unpitched musicians should perform

similarly to the nonmusicians. To recap, the following research questions motivated our study:

1. Did pitched musicians outperform unpitched musicians and nonmusicians on lexical tone discrimination and sequence recall?
2. Did unpitched musicians perform similarly to nonmusicians on lexical tone discrimination and sequence recall?

Method

PARTICIPANTS

We recruited 61 Cantonese listeners (33 males and 27 females) aged from 19 to 33 years ($M = 22.8$ years; $SD = 2.61$ years) in Hong Kong via email and posters. All participants provided informed consent and completed a language and music background questionnaire (Choi, 2021, 2022a; Choi & Lai, 2023). According to self-report, all Cantonese listeners spoke Cantonese as first language, had normal hearing, and had no previous Thai learning experience and absolute pitch. We classified the participants into three groups, i.e., pitched musicians, unpitched musicians, and nonmusicians. Based on the criteria used in previous studies (Choi, 2020, 2022b; Tong et al., 2018), the pitched musicians had at least seven years of continuous piano and/or violin training, less than two years of unpitched percussion training, and could play their instruments at the time of testing. The unpitched musicians had at least seven years of continuous unpitched percussion training, less than two years of pitched music training, and could play their instruments at the time of testing. The nonmusicians had less than two years of music training, no music training in the past five years, and could not play any musical instrument at the time of testing. We excluded three unpitched percussionists from the study due to excess pitched instrument learning experience. The final sample contained 20 pitched musicians (10 males, 10 females), 18 unpitched musicians (11 males, 7 females) and 20 nonmusicians (10 males, 10 females).

The mean ages of the pitched musicians, unpitched musicians and nonmusicians were 21.95 years ($SD = 1.79$ year), 23.72 years ($SD = 3.17$ years), and 23.00 years ($SD = 1.97$ year), respectively. On average, the pitched musicians and unpitched musicians had 10.35 years ($SD = 3.42$ years) and 9.50 years of music training ($SD = 2.81$ years), respectively, while the nonmusicians only had 1.30 years of music training ($SD = 0.45$ year) if at all. Tables 1 and 2 summarize the instruments learnt by the pitched and unpitched musicians.

TABLE 1. *Music Background of the Pitched Musicians*

Participants	Years of music training	First instrument	Second instrument	Third instrument
1	8	Piano		
2	7	Piano		
3	8	Piano	Cello	Guitar
4	11	Violin	Piano	
5	13	Piano		
6	7	Piano		
7	7	Piano		
8	8	Piano		
9	11	Piano		
10	8	Piano	Cello	Organ
11	8	Piano		
12	16	Violin	Piano	Viola
13	10	Piano	Flute	Pitched percussion
14	12	Violin		
15	10	Piano	Oboe	
16	9	Piano	Guitar	
17	17	Piano		
18	7	Violin		
19	18	Piano	Organ	
20	12	Piano		

TABLE 2. *Music Background of the Unpitched Musicians*

Participants	Years of music training	First instrument	Second instrument	Third instrument
1	12.5	Snare drum	Guitar*	Drum set
2	12	Drum set	Piano*	
3	7	Drum set		
4	7	Drum set		
5	8	Bass drum		
6	8	Drum set		
7	10	Drum set		
8	18	Drum set	Keyboard*	
9	7	Drum set	Cajon*	Guitar*
10	13	Drum set	Piano*	
11	10	Drum set		
12	8	Drum set		
13	10	Drum set	Piano*	
14	12	Drum set		
15	7	Drum set		
16	8	Drum set		
17	11	Drum set		
18	7	Drum set	Piano*	

Note. * less than two years.

PROCEDURE

The experiment took place in a sound booth at The University of Hong Kong or in a music studio. Participants completed the Thai tone discrimination and sequence recall tasks under the supervision of the second or the third author. We ran the experiment on

a laptop with E-prime 3.0. Throughout the experiment, participants heard all audio stimuli via Sennheiser HD280 PRO headphones. The entire experiment lasted for about 40 minutes. We had obtained ethical approval from the Faculty Research Ethics Committee of the Faculty of Education at The University of Hong Kong.

LEXICAL TONE DISCRIMINATION TASK

Stimuli

Thai language has five lexical tones, including three level tones (high, mid, and low) and two contour tones (falling and rising), e.g., ป่า [paa – mid (M)] *throw*, ป่า [pàa – low (L)] *forest*, ป่า [pâa – falling (F)] *aunt*, ป่า [pâa – high (H)] *father*, ป่า [pǎa – rising (R)] *dad* (Cooke, 1963; Winskel, 2011). This yielded 10 Thai tone contrasts; that is, M-L, M-F, M-H, M-R, L-F, L-H, L-R, F-H, F-R, H-R.

Two native Thai speakers (one male and one female) recorded the stimuli with a Shure SM58 microphone at a sampling rate of 48k Hz. The two speakers naturally produced the five Thai tones embedded in different segments, e.g., /goht/, /glaa/, /yoo/, /miia/, and /meet/. To preserve the naturalness of the stimuli, we did not acoustically manipulate them (e.g., Choi, 2020; Cooper & Wang, 2012; Nan et al., 2018). The natural duration of the stimuli used in our study ranged from 420 ms to 1,394 ms. Although the decibels relative to full scale (RMS DB FS) was not equated between all the stimuli, our acoustic-behavioural analysis showed that the accuracies (of pitched musicians, unpitched musicians, non-musicians, and all groups combined) were not associated with it, $ps = .749, .587, .774, \text{ and } .948$. Before and during the experiment, participants could adjust the volume of the laptop. All participants reported that they could hear the stimuli clearly at a comfortable volume. The stimuli are available in Open Science Framework (<https://osf.io/2p6hb>).

Stimuli Presentation

We adopted the AXB paradigm (Choi, 2020). The task contained 10 pairs of Thai tone contrasts (i.e., M-L, M-F, M-H, M-R, L-F, L-H, L-R, F-H, F-R, and H-R). On each trial, listeners heard three syllables (e.g., /goht-M/ /goht-R/ /goht-R/) at 800 ms apart. Then, they pressed the associated keys to indicate if the first or the third syllable carried the same lexical tone as the second one did. AB and X were produced by speakers of different genders to prevent the participants' reliance on simple acoustic comparisons. There were eight trials for each Thai tone contrast and the total number of trials was 80 (10 pairs of Thai tone contrasts x 8 trials). Prior to the task, participants were explicitly requested to attend to

the lexical tones, but not duration, intensity, and voice. The task began with four practice trials with feedback. In each experimental trial, the accuracy was recorded and no feedback was provided. The sample-specific internal consistency was satisfactory (Cronbach's $\alpha = .78$).

SEQUENCE RECALL TASK

Stimuli

The same native Thai speakers as above recorded a vowel contrast, i.e., /rɛ-M/-/ru-M/, and two lexical tone contrasts, i.e., /phɪ-M/-/phɪ-H/ and /pluuk-F/-/pluuk-R/. The vowel contrast contained the same mid tone. The first tone contrast contained two level tones (mid and high) whereas the second one contained two contour tones (falling and rising). The stimuli are available in Open Science Framework (<https://osf.io/2p6hb>).

Stimuli Presentation

Adapted from the stress sequence recall task (Dupoux et al., 2010), this task assessed the ability to represent and store Thai tones in memory. The task contained the vowel context, the level tone context (/phɪ-M/-/phɪ-H/), and the contour tone context (pluuk-F/-/pluuk-R/). Each context began with a familiarization phase in which participants listened to two items (as many times as desired) and associated them with their corresponding keys [1] and [2]. Then, there were five identification trials with feedback. In the sequence recall phase, participants listened to a sequence of syllables (600 ms apart) with varying length (two to six) and reproduced the sequence by pressing the associated keys, e.g., pressing [1] [2] [2] [1] for /phɪ-M/-/phɪ-H/-/phɪ-H/-/phɪ-M/. On each trial, a response was considered correct only if it fully matched the sequence presented. In each context, there were six trials for each sequence length. The total number of trials was 60. The sample-specific internal consistencies were high in the vowel context (Cronbach's $\alpha = .80$) and very high in the level tone (Cronbach's $\alpha = .90$) and contour tone contexts (Cronbach's $\alpha = .91$).

WORKING MEMORY TASK

We adapted the backward digit span task from the Wechsler Adult Intelligence Scale (Fourth Edition, Hong Kong; Wechsler, 2014). WAIS-IV(HK) was specifically developed for the Cantonese speaking population in Hong Kong aged between 16 and 90 years. The task was administered in Cantonese. On each trial, participants listened to a sequence of digits (two to eight) verbally produced by the experimenter in Cantonese, and recalled the sequence in reverse order. The

sequence length increased by one after correct responses of two sequences at the same length until the participant failed to recall both sequences. The maximum sequence length recalled by each participant was collected. The sample-specific internal consistency was fair (Cronbach's $\alpha = .65$).

NONVERBAL INTELLIGENCE TASK

We conducted the Raven's 2 Progressive Matrices to assess nonverbal intelligence (Raven et al., 2018). In each of the 24 questions, participants chose the appropriate picture to complete a visual pattern. One point was given for each correct answer. The total score was collected. The sample-specific internal consistency was satisfactory (Cronbach's $\alpha = .73$).

Results

PRELIMINARY ANALYSIS

We conducted two sets of one-way ANOVAs to examine whether the three groups differed in working memory and nonverbal intelligence. The main effect of group was not significant for working memory, $p = .402$, and nonverbal intelligence, $p = .095$. These results indicate that the three groups matched on working memory and nonverbal intelligence.

To be empirically stringent, we still controlled for working memory and nonverbal intelligence in the main analysis below. The same set of main analysis without the two control variables yielded entirely consistent results. For both sets of analysis, we have uploaded the annotated.jasp files including data and input options to Open Science Framework (<https://osf.io/2p6hb>).

MAIN ANALYSIS

Our two research questions were: 1) whether pitched musicians outperformed unpitched musicians and nonmusicians on Thai tone discrimination and sequence recall; and 2) whether unpitched musicians performed similarly to nonmusicians in these tasks. As reasoned in Appendix A, we will adopt the Bayesian approach to address these research questions.

Lexical Tone Discrimination

We conducted a Bayesian two-way ANCOVA on mean accuracy with JASP 0.17.3 (JASP Team, 2023). The within-subject factor was lexical tone contrast (M-L, M-F, M-H, M-R, L-F, L-H, L-R, F-H, F-R, and H-R), the between-subjects factor was group (pitched musicians, unpitched musicians, and nonmusicians), and the covariates were working memory and nonverbal

TABLE 3. Model Comparison Relative to the Best-fit Model of Discrimination Accuracy

Models	P(M)	P(M data)	BF _M	BF ₀₁	error %
Tone + Group	0.050	0.433	14.537	1.000	
Tone	0.050	0.260	6.661	1.670	2.380
Tone + Group + WM	0.050	0.077	1.590	5.613	3.945
Tone + Group + IQ	0.050	0.073	1.505	5.905	2.991
Tone + WM	0.050	0.067	1.362	6.479	33.803
Tone + IQ	0.050	0.053	1.069	8.135	2.608
Tone + Group + WM + IQ	0.050	0.020	0.381	22.072	3.091
Tone + WM + IQ	0.050	0.014	0.276	30.314	2.843
Tone + Group + Tone × Group	0.050	0.002	0.030	270.809	2.647
Tone + Group + IQ + Tone × Group	0.050	2.720×10 ⁻⁴	0.005	1593.720	2.733
Null	0.050	1.094×10 ⁻³⁸	2.079×10 ⁻³⁷	3.962×10 ⁻³⁷	2.365

Note. Showing the best 10 models and the null model, in ascending order of BF₀₁. Tone = tone contrast; WM = working memory; IQ = nonverbal intelligence.

TABLE 4. Post hoc Comparisons of the Groups on Discrimination Accuracy

		Prior Odds	Posterior Odds	BF _{10, U}	error %
Pitched	Unpitched	0.587	2.623	4.466	0.005
	Nonmusicians	0.587	111.288	189.458	1.420×10 ⁻⁴
Unpitched	Nonmusicians	0.587	0.109	0.186	0.087

Note. The posterior odds were corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall et al., 1997). Individual comparisons were based on the default *t*-test with a Cauchy (0, $r = 1/\sqrt{2}$) prior. BF_{10, U} denotes uncorrected BF₁₀. Pitched = pitch musicians; Unpitched = unpitched musicians.

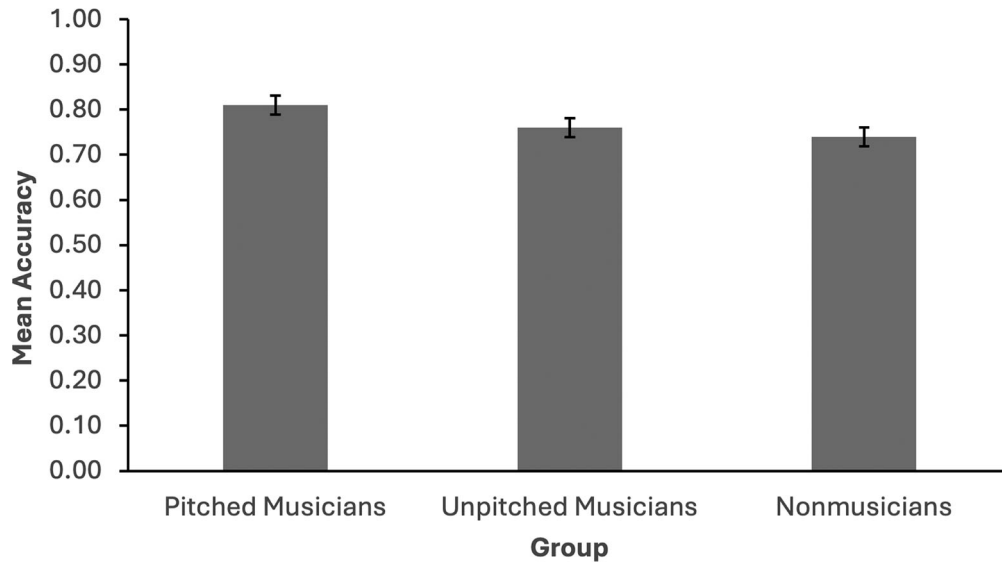


FIGURE 1. Mean accuracies of pitched musicians, unpitched musicians, and nonmusicians collapsed across tone pairs in the Thai tone discrimination task. Note. Error bars denote 95% confidence intervals.

intelligence. In total, there were 20 models. We conducted model comparisons relative to the best-fit model (see Table 3). Bayes factors indicated that the data best

supported the model (Tone + Group) with significant main effects of tone contrast and group. Specifically, BF₀₁ suggested that the null model and the other models

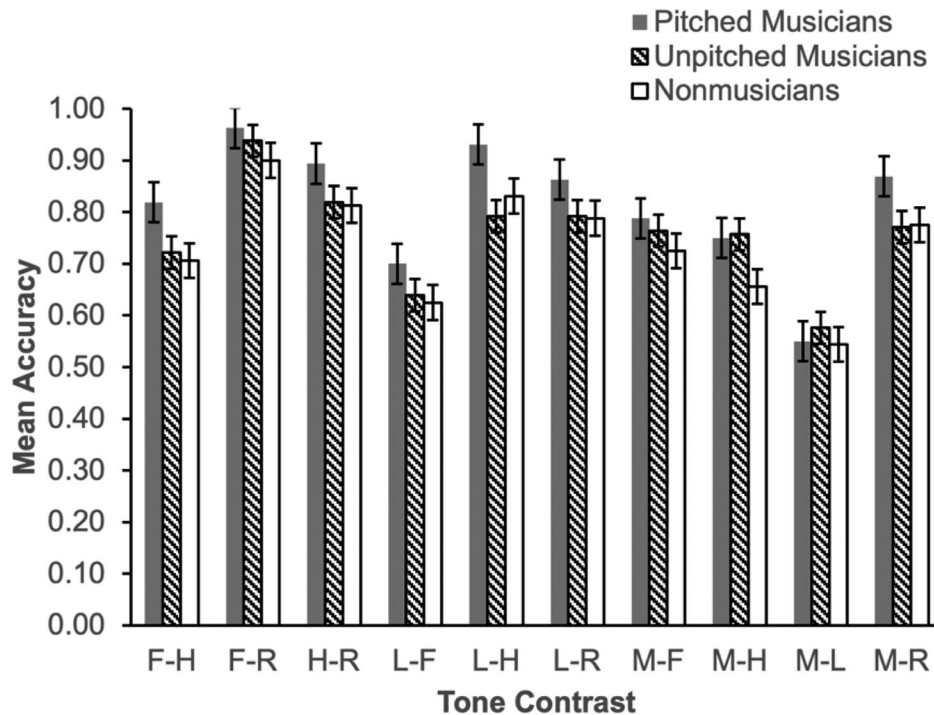


FIGURE 2. Mean accuracies of pitched musicians, unpitched musicians, and nonmusicians on individual tone pairs in the Thai tone discrimination task. Note. Error bars denote 95% confidence intervals.

were 3.96×10^{37} and $1.67\text{--}1.10 \times 10^{39}$ times less likely than the best-fit model. Model comparisons relative to the null model yielded consistent results (see Table A2). Our central focus was the main effect of group. According to post hoc comparisons, the pitched musicians outperformed the unpitched musicians and the nonmusicians with moderate ($BF_{10, U} = 4.47$) and extreme evidence ($BF_{10, U} = 189.46$) respectively (see Table 4). However, the unpitched musicians and the nonmusicians performed similarly, with moderate evidence ($BF_{10, U} = 0.19$).

Sequence Recall

We conducted a Bayesian two-way ANCOVA on mean accuracy with context (vowel, level tone, and contour tone) as the within-subject factor, group (pitched musicians, unpitched musicians, and nonmusicians) as the between-subjects factor, and working memory and nonverbal intelligence as the co-variates. In total, there were 20 models. We conducted model comparisons relative to the best-fit model (see Table 5). Bayes factor indicated that the data best supported the model (Context + Group + IQ + Context \times Group) with significant main effects of context, group, nonverbal intelligence, and the interaction between context and group. Specifically,

BF_{01} suggested that the null model and the other models were 5.40×10^6 and $1.28\text{--}1.68 \times 10^7$ times less likely than the best-fit model. Model comparisons relative to the null model yielded consistent results (see Table A3). Our central focus was the interaction between context and group.

To unpack the interaction between context and group, we conducted a Bayesian one-way ANCOVA in each context with mean accuracy as the dependent variable, group (pitched musicians, unpitched musicians, and nonmusicians) as the between-subjects factor, and working memory and nonverbal intelligence as the covariates. In the vowel context, the best-fit model was the model (Group + IQ) with significant main effects of group and nonverbal intelligence. Specifically, BF_{01} indicated that the null model and the other models were 17.10 and 1.04–40.02 times less likely than the best-fit model (see Table 6; see also Table A4 for BF_{10}). For the main effect of group, post hoc comparisons indicated that the pitched musicians and unpitched musicians outperformed the non musicians with moderate evidence, $BF_{10, U} = 7.68$ and 7.92. However, the pitched musicians and unpitched musicians performed similarly with moderate evidence, $BF_{10, U} = 0.326$ (see Table 7).

TABLE 5. Model Comparison Relative to the Best-fit Model of Sequence Recall Accuracy

Models	P(M)	P(M data)	BF _M	BF ₀₁	error %
Context + Group + IQ + Context× Group	0.050	0.404	12.898	1.000	
Context + Group + Context× Group	0.050	0.316	8.786	1.279	4.295
Context + Group + WM + IQ + Context× Group	0.050	0.161	3.650	2.509	2.278
Context + Group + WM + Context× Group	0.050	0.089	1.860	4.535	2.301
Context + Group + IQ	0.050	0.012	0.234	33.250	4.796
Context + Group	0.050	0.009	0.172	45.145	2.469
Context + Group + WM + IQ	0.050	0.004	0.084	91.924	1.908
Context + Group + WM	0.050	0.003	0.048	160.366	2.061
Context + IQ	0.050	7.579×10 ⁻⁴	0.014	533.525	28.846
Context + WM + IQ	0.050	1.892×10 ⁻⁴	0.004	2136.758	1.730
Null	0.050	7.485×10 ⁻⁸	1.422×10 ⁻⁶	5.402×10 ⁶	1.440

Note. Showing the best 10 models and the null model, in ascending order of BF₀₁. Tone = tone contrast; WM = working memory; IQ = nonverbal intelligence.

TABLE 6. Model Comparison Relative to the Best-fit Model of Vowel Sequence Recall Accuracy

Models	P(M)	P(M data)	BF _M	BF ₀₁	error %
Group + IQ	0.125	0.339	3.590	1.000	
Group	0.125	0.327	3.409	1.035	1.280
Group + WM + IQ	0.125	0.131	1.058	2.583	2.046
Group + WM	0.125	0.100	0.777	3.392	1.662
IQ	0.125	0.052	0.386	6.485	1.280
WM + IQ	0.125	0.022	0.155	15.620	1.280
Null model	0.125	0.020	0.142	17.099	1.280
WM	0.125	0.008	0.060	40.016	1.280

Note. WM = working memory; IQ = nonverbal intelligence.

TABLE 7. Post hoc Comparisons of the Groups on Vowel Sequence Recall Accuracy

		Prior Odds	Posterior Odds	BF _{10, U}	error %
Pitched	Unpitched	0.587	0.192	0.326	0.004
	Nonmusicians	0.587	4.513	7.683	1.752×10 ⁻⁶
Unpitched	Nonmusicians	0.587	4.653	7.921	2.429×10 ⁻⁶

Note. Pitched = pitched musicians; Unpitched = unpitched musicians.

TABLE 8. Model Comparison Relative to the Best-fit Model of Level Tone Sequence Recall Accuracy

Models	P(M)	P(M data)	BF _M	BF ₀₁	error %
Group + IQ	0.125	0.398	4.633	1.000	
Group	0.125	0.361	3.951	1.104	0.887
Group + WM + IQ	0.125	0.142	1.161	2.799	1.518
Group + WM	0.125	0.098	0.757	4.081	2.997
IQ	0.125	6.729×10 ⁻⁴	0.005	591.904	0.886
WM + IQ	0.125	2.150×10 ⁻⁴	0.002	1852.747	0.886
Null model	0.125	1.232×10 ⁻⁴	8.624×10 ⁻⁴	3233.060	0.886
WM	0.125	3.734×10 ⁻⁵	2.614×10 ⁻⁴	10666.525	0.886

Note. WM = working memory; IQ = nonverbal intelligence.

TABLE 9. Post hoc Comparisons of the Groups on Level Tone Sequence Recall Accuracy

		Prior Odds	Posterior Odds	BF _{10, U}	error %
Pitched	Unpitched	0.587	1.800	3.064	0.009
	Nonmusicians	0.587	7112.973	12109.228	5.620×10 ⁻¹⁰
Unpitched	Nonmusicians	0.587	2.791	4.751	8.185×10 ⁻⁷

Note. Pitched = pitched musicians; Unpitched = unpitched musicians.

TABLE 10. Model Comparison Relative to the Null and Best-fit Model of Contour Tone Sequence Recall Accuracy

Models	P(M)	P(M data)	BF _M	BF ₀₁	error %
Null model	0.125	0.283	2.762	1.000	
IQ	0.125	0.229	2.074	1.238	0.001
Group	0.125	0.163	1.368	1.731	0.025
WM + IQ	0.125	0.086	0.658	3.291	0.006
Group + IQ	0.125	0.083	0.634	3.407	1.037
WM	0.125	0.076	0.572	3.746	0.003
Group + WM	0.125	0.045	0.330	6.276	1.139
Group + WM + IQ	0.125	0.035	0.257	7.976	2.480

Note. WM = working memory; IQ = nonverbal intelligence.

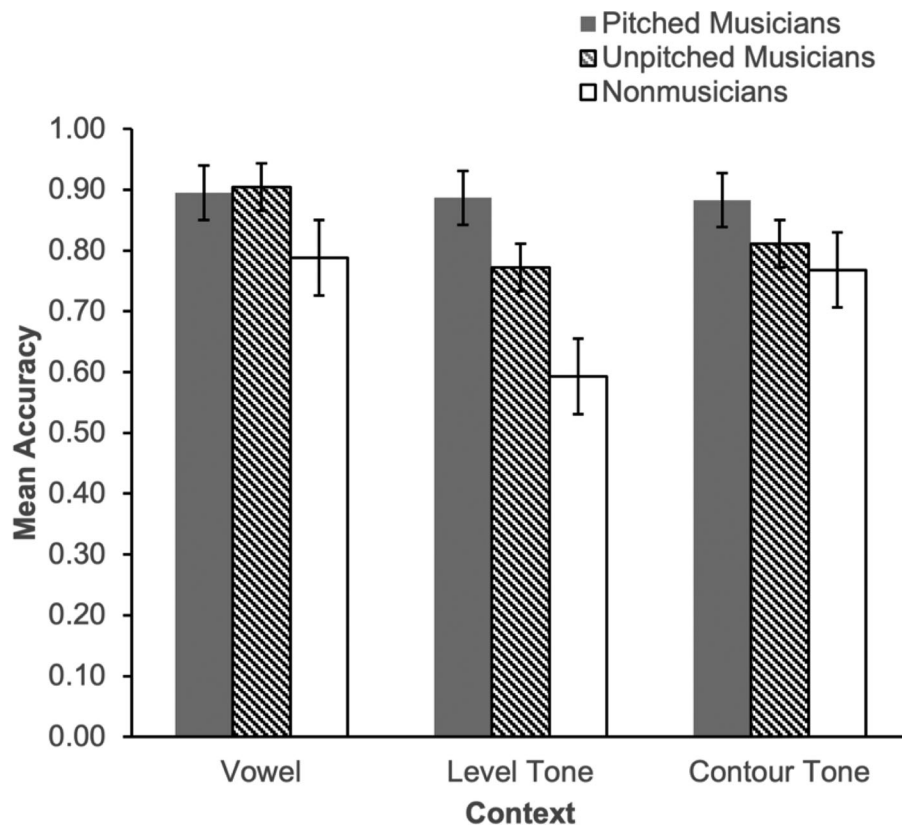


FIGURE 3. Mean accuracies of pitched musicians, unpitched musicians, and nonmusicians in the sequence recall task. Note. Error bars denote 95% confidence intervals.

In the level tone context, the best-fit model was the model (Group + IQ) with significant main effects of group and nonverbal intelligence. Specifically, BF_{01} suggested that the null model and the other models were 3233.06 and 1.10–10666.53 times less likely than the best-fit model (see Table 8; see also Table A5 for BF_{10}). For the main effect of group, post hoc comparisons showed that the pitched musicians outperformed the unpitched musicians and the nonmusicians with moderate evidence ($BF_{10, U} = 3.06$) and extreme evidence ($BF_{10, U} = 12109.23$) respectively. Moreover, the unpitched musicians outperformed the nonmusicians with moderate evidence, $BF_{10, U} = 4.75$ (see Table 9).

In the contour tone context, the model comparisons favored the null model. Specifically, BF_{01} suggested that the other models were 1.24–7.98 times less likely than the null model (see Table 10).

To supplement the above analysis, we also unpacked the interaction between context and group at each level of group. The results are summarized in Appendix B.

Discussion

The present study examined: 1) whether pitched musicians outperformed unpitched musicians and nonmusicians on Thai tone discrimination and sequence recall; and 2) whether unpitched musicians performed similarly to nonmusicians in these tasks. For lexical tone discrimination, the pitched musicians outdid the unpitched musicians and the nonmusicians. Moreover, the unpitched musicians performed similarly to the nonmusicians. Regarding lexical tone sequence recall, the pitched musicians outperformed the unpitched musicians and the nonmusicians in the level tone context. Moreover, the unpitched musicians also outdid the nonmusicians on level tone sequence recall. However, the three groups performed similarly in the contour tone context.

The principal finding is the unique advantage of pitched musicians in lexical tone discrimination. Consistent with our hypothesis, the pitched musicians discriminated Thai tones more accurately than the unpitched musicians and the nonmusicians. Importantly, the unpitched musicians did not outperform the nonmusicians. This has reflected that the musical advantage in lexical tone discrimination is unique to pitched musicians. Furthermore, the above results apply to all the Thai tone contrasts. On the one hand, our results offer correlational evidence for the Precision element of OPERA (Patel, 2011, 2014). For music-to-language transfer to occur in lexical tone perception, the Precision element assumes that music must entail more precise pitch processing than speech does. Pitched musical instruments

have a high demand on pitch processing, so our pitched musicians exhibited an advantage in lexical tone discrimination. By contrast, unpitched musical instruments have minimal demand on pitch processing. Thus, our unpitched musicians showed no advantage in lexical tone discrimination relative to the nonmusicians. Without considering the heterogeneity of musical instruments, previous lexical tone perception studies suggested that music-to-language transfer occurred with music training (e.g., Choi, 2020; Lee et al., 2014; Zheng & Samuel, 2018). Extending these studies, our study further suggests that music-to-language transfer in lexical tone discrimination occurs only when the music training has a high processing demand on pitch, consistent with OPERA's notion (Patel, 2011, 2014).

On the other hand, it is also possible that our pitched musicians had higher pre-existing musical pitch sensitivity than the unpitched musicians and the nonmusicians even well before music training (see Schellenberg, 2015). From an empirical perspective, we have provided correlational evidence to support the feasibility of an RCT study that compares the effectiveness of pitched versus unpitched instrumental training in lexical tone perception. If the pitched instrumental training group shows larger gains than the unpitched instrumental training and waitlist control groups, it would provide OPERA with strong causal evidence of the Precision element.

Consistent with our hypothesis, the pitched musicians outperformed the unpitched musicians and the nonmusicians on level tone sequence recall. Unexpectedly, even though the unpitched musicians underperformed the pitched musicians, the former still outdid the nonmusicians. Unlike the perceptual-based discrimination task, the sequence recall task entails both perception and sequence recall. Participants not only had to perceive the lexical tones, but also remember and reproduce the sequences. Music training typically requires musicians to remember and reproduce melodic and rhythmic sequences (Bigand & Tillmann, 2022; Talamini et al., 2017). This may translate to the cognitive ability to remember and reproduce sequences in the sequence recall task, benefiting the pitched and the unpitched musicians. On top of this cognitive enhancement on sequence recall which both groups enjoyed, the pitched musicians had better level tone discrimination. This further enabled the pitched musicians to encode the level tone sequences more efficiently than the unpitched musicians. Thus, the pitched musicians exhibited the largest musical advantage, and the unpitched musicians enjoyed a small musical advantage relative to the nonmusicians. Critically, our working memory measure (backward digit span) did not correlate significantly with level tone

sequence recall ($p = .578$). Possibly, if we had controlled for a working memory measure that is more closely related to sequence recall (e.g., forward span), or even other cognitive measures (Benz et al., 2016; Zuk et al., 2014), the unpitched musicians might have performed similarly to the nonmusicians.

Inconsistent with our hypothesis, there was no musical advantage in contour tone sequence recall. At first glance, this finding paralleled a previous study on the selectivity of musical advantage in English musicians (Choi, 2020). However, there are two major discrepancies between the current results and the earlier work. First, English musicians outdid English nonmusicians on recalling contour but not level tone sequences, but Cantonese musicians showed the opposite pattern. This might be due to differences in their first language (English and Cantonese), the stimuli (Cantonese and Thai), and the focus of their music training. The most important discrepancy is that there are different underlying reasons for the observed selectivity of musical advantage between the two studies. In the previous study, English musicians performed worse on level than contour tone sequences, implying that musicianship did not benefit level tone sequence recall much, if at all (Choi, 2020; cf. Schellenberg, 2015). Contrastively, the Cantonese musicians recalled both sequences with similar accuracy, but the nonmusicians performed better on contour than level tone sequences. Unlike in the previous study, the loss of musical advantage was because the nonmusicians caught up with the musicians in recalling contour tone sequences. Although the nonmusicians discriminated contour tones less accurately than the pitched musicians in the AXB task, they might have developed a compensatory strategy to achieve the same performance as the pitched musicians in sequence recall. Future studies are needed to identify their compensatory strategy and why the strategy did not apply to level tones. From a theoretical perspective, the collective findings suggest that the Precision element may not apply to the recall of every lexical tone sequence. Moreover, there could be different reasons giving rise to the selectivity of musical advantage. For example, it could be musicianship not being associated with enhanced perception of some lexical tone contrasts (Choi, 2020), or nonmusicians catching up with musicians.

An incidental finding is that the pitched and unpitched musicians recalled vowel sequences more accurately than the nonmusicians. Acoustically, / ϵ / and / u / differ in formant frequencies and duration. Loui et al. (2011) proposed that musical pitch sensitivity underlies the processing of first and second formant frequencies. Moreover, Tierney and Kraus (2014) suggested that rhythmic sensitivity aids the perception of fine timing

details in speech. Indeed, our pitched and unpitched musicians showed an advantage, reflecting the roles of pitch and rhythm in vocalic perception. Furthermore, the equal performance of the pitched and the unpitched musicians further suggests that pitch and rhythm may play equally important roles. We encourage researchers to conduct RCT studies on the effectiveness of pitch and rhythm training on vowel perception.

We have identified several avenues for future research. First, our correlational evidence suggests that it is feasible to conduct an RCT study on the causal effect of pitched and unpitched instrumental training on lexical tone perception. Second, our pitched musician group contained pianists and violinists. While pianists control the pitch over an equal-tempered scale, violinists need to monitor and regulate a constantly changing pitch (with finger placement, bow pressure, and string tension) (Shahin et al., 2003). Compared with level tones, contour tones have a dynamic F0 contour (Choi, 2020). Thus, researchers can examine whether violinists exhibit a larger musical advantage in contour tones compared with pianists. Third, musicianship is a multifaceted concept that is well beyond musical instrument. Cross-culturally, Western music pays more attention to steady-state pitch sequences (Leech-Wilkinson, 2006) whereas Chinese music has a high occurrence of gliding pitches (Shi, 2016). Researchers can examine the differential effects of Western and Chinese music training on level and contour tone perception. Lastly, researchers have started to distinguish between professional musicians and amateur musicians (Oechslin et al., 2013; Rogenmoser et al., 2017; Schneider et al., 2002) and enduring musicians and former musicians (Toh et al., 2023). These fine distinctions prompt nuanced investigations of music-to-language transfer.

To conclude, pitched musicians showed a unique musical advantage in lexical tone discrimination. In level tone sequence recall, although both pitched and unpitched musicians outdid the nonmusicians, the pitched musicians had the largest musical advantage. Taken together, the pitch processing demand of musical instrument may matter for music-to-language transfer in lexical tone discrimination and level tone sequence recall. From a theoretical perspective, this offers correlational support for the Precision element of OPERA (Patel, 2011, 2014). From a practical perspective, there is a trend of utilizing music training to enhance speech perception (e.g., Kraus & Chandrasekaran, 2010; Nan et al., 2018). Our study further suggests that the choice of musical instrument may matter, at least for lexical tone perception. In addition to musical instrument, there are many more ways to decipher musicianship (e.g., Oechslin et al., 2013; Rogenmoser et al., 2017; Toh

et al., 2023). Deciphering musicianship can certainly advance the theoretical understanding of music-to-language transfer and its practical application.

Author Note

We have no known conflict of interest to disclose. This study was based, in part, on the undergraduate final year projects of the second and third authors. We appreciate

Kusol Im-Erbsin and Ratana-U-Bol for their assistance in stimuli development. This work was supported by The University of Hong Kong (Seed Fund for Basic Research for New Staff [202107185043] and Project-Based Research Funding [supported by Start-up Fund]).

Correspondence concerning this article should be addressed to Professor William Choi, Room 757, Meng Wah Complex, The University of Hong Kong, Pokfulam, Hong Kong. E-mail: willchoi@hku.hk

References

- ALEXANDER, J. A., WANG, P. C. M., & BRADLOW, A. R. (2005). *Lexical tone perception in musicians and non-musicians* [Paper presentation]. 9th European Conference on Speech Communication and Technology, Lisbon.
- BENZ, S., SELLARO, R., HOMMEL, B., & COLZATO, L. S. (2016). Music makes the world go round: The impact of musical training on non-musical cognitive functions—A review. *Frontiers in Psychology*, 6, 2023. <https://doi.org/10.3389/fpsyg.2015.02023>
- BIGAND, E., & TILLMANN, B. (2022). Near and far transfer: Is music special? *Memory and Cognition*, 50(2), 339–347. <https://doi.org/10.3758/s13421-021-01226-6>
- BURNHAM, D., BROOKER, R., & REID, A. (2015). The effects of absolute pitch ability and musical training on lexical tone perception. *Psychology of Music*, 43(6), 881–897. <https://doi.org/10.1177/0305735614546359>
- BURNHAM, D., KASISOPA, B., REID, A., LUKSANEYANAWIN, S., LACERDA, F., ATTINA, V., ET AL. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics*, 36(6), 1459–1491. <https://doi.org/10.1017/S0142716414000496>
- CAMERON, D. J., & GRAHN, J. A. (2014). Enhanced timing abilities in percussionists generalize to rhythms without a musical beat. *Frontiers in Human Neuroscience*, 8, 1003. <https://doi.org/10.3389/fnhum.2014.01003>
- CHEN, J., BEST, C. T., ANTONIOU, M., & KASISOPA, B. (2019). Cognitive factors in perception of Thai tones by naïve Mandarin listeners. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, (pp. 1684–1688). Australasian Speech Science and Technology Association Inc.
- CHEN, J. Q., ANTONIOU, M., & BEST, C. T. (2023). Phonological and phonetic contributions to perception of non-native lexical tones by tone language listeners: Effects of memory load and stimulus variability. *Journal of Phonetics*, 96, 101199. <https://doi.org/10.1016/j.wocn.2022.101199>
- CHEN, S., ZHU, Y., WAYLAND, R., & YANG, Y. (2020). How musical experience affects tone perception efficiency by musicians of tonal and non-tonal speakers? *PLOS ONE*, 15(5), e0232514. <https://doi.org/10.1371/journal.pone.0232514>
- CHOI, W. (2020). The selectivity of musical advantage: Musicians exhibit perceptual advantage for some but not all Cantonese tones. *Music Perception*, 37(5), 423–434. <https://doi.org/10.1525/MP.2020.37.5.423>
- CHOI, W. (2021). Musicianship influences language effect on musical pitch perception. *Frontiers in Psychology*, 12, 712753. <https://doi.org/10.3389/fpsyg.2021.712753>
- CHOI, W. (2022a). What is ‘music’ in music-to-language transfer? Musical ability but not musicianship supports Cantonese listeners’ English stress perception. *Journal of Speech, Language, and Hearing Research*, 65, 4047–4059. https://doi.org/10.1044/2022_JSLHR-22-00175
- CHOI, W. (2022b). Towards a native OPERA hypothesis: Musicianship and English stress perception. *Language and Speech*, 65(3), 697–712. <https://doi.org/10.1177/00238309211049458>
- CHOI, W., & LAI, V. K. W. (2023). Does musicianship influence the perceptual integrality of tones and segmental information? *Journal of the Acoustical Society of America*, 154(2), 852–862. <https://doi.org/10.1121/100.0020579>
- CICCHINI, G. M., ARRIGHI, R., CECCHETTI, L., GIUSTI, M., & BURR, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience*, 32(3), 1056–1060. <https://doi.org/10.1523/JNEUROSCI.3411-11.2012>
- COFFEY, E. B. J., MOGILEVER, N. B., & ZATORRE, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, 352, 49–69. <https://doi.org/10.1016/j.heares.2017.02.006>
- COOKE, J. R. (1963). The vowels and tones of standard Thai: Acoustical measurements and experiments. Arthur S. Abramson. *American Anthropologist*, 65, 1406–1407.
- COOPER, A., & WANG, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *Journal of the Acoustical Society of America*, 131(6), 4756–4769. <https://doi.org/10.1121/1.4714355>
- DELOGU, F., LAMPIS, G., & BELARDINELLI, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, 22(1), 46–61. <https://doi.org/10.1080/09541440802708136>

- DIENES, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. <https://doi.org/10.3389/fpsyg.2014.00781>
- DIENES, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72, 78–89. <https://doi.org/10.1016/j.jmp.2015.10.003>
- DUPOUX, E., PEPERKAMP, S., & SEBASTIÁN-GALLÉS, N. (2010). Limits on bilingualism revisited: Stress ‘deafness’ in simultaneous French-Spanish bilinguals. *Cognition*, 114(2), 266–275. <https://doi.org/10.1016/j.cognition.2009.10.001>
- GORDON, R. L., FEHD, H. M., & McCANDLISS, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. *Frontiers in Psychology*, 6, 1777. <https://doi.org/10.3389/fpsyg.2015.01777>
- HALLÉ, P. A., CHANG, Y.-C., & BEST, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32(3), 395–421. [https://doi.org/10.1016/S0095-4470\(03\)00016-0](https://doi.org/10.1016/S0095-4470(03)00016-0)
- HENNESSY, S., MACK, W. J., & HABIBI, A. (2022). Speech-in-noise perception in musicians and non-musicians: A multi-level meta-analysis. *Hearing Research*, 416, 108442. <https://doi.org/10.1016/j.heares.2022.108442>
- JASP TEAM (2023). *JASP (Version 0.17.3)* [Computer software].
- JAYNES, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511790423>
- JOHNSON, K., FLEMMING, E., & WRIGHT, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language (Baltimore)*, 69(3), 505–528. <https://doi.org/10.2307/416697>
- KIM, H., & TREMBLAY, A. (2021). Korean listeners’ processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. *Journal of Phonetics*, 87, 101059. <https://doi.org/10.1016/j.wocn.2021.101059>
- KRAUS, N., & CHANDRASEKARAN, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599–605. <https://doi.org/10.1038/nrn2882>
- LEE, C. Y., LEKICH, A., & ZHANG, Y. (2014). Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America*, 135(3), 1607–1615. <https://doi.org/10.1121/1.4864473>
- LEE, M. D., & WAGENMAKERS, E.-J. (2013). *Bayesian cognitive modelling: A practical course*. Cambridge University Press.
- LEECH-WILKINSON, D. (2006). Portamento and musical meaning. *Journal of Musicological Research*, 25(3–4), 233–261. <https://doi.org/10.1080/01411890600859412>
- 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>
- LY, A., ETZ, A., MARSMAN, M., & WAGENMAKERS, E.-J. (2019). Replication Bayes factors from evidence updating. *Behavior Research Methods*, 51(6), 2498–2508. <https://doi.org/10.3758/s13428-018-1092-x>
- MAILLARD, E., JOYAL, M., MURRAY, M. M., & TREMBLAY, P. (2023). Are musical activities associated with enhanced speech perception in noise in adults? A systematic review and meta-analysis. *Current Research in Neurobiology*, 4, 100083. <https://doi.org/10.1016/j.crneur.2023.100083>
- MORENO, S., MARQUES, C., SANTOS, A., SANTOS, M., CASTRO, S. L., & BESSON, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, 19(3), 712–723. <https://doi.org/10.1093/cercor/bhn120>
- NAN, Y., LIU, L., GEISER, E., SHU, H., GONG, C. C., DONG, Q., ET AL. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences*, 115(28), E6630–E6639.
- OECHSLIN, M. S., VAN DE VILLE, D., LAZEYRAS, F., HAUERT, C.-A., & JAMES, C. E. (2013). Degree of musical expertise modulates higher order brain functioning. *Cerebral Cortex*, 23(9), 2213–2224. <https://doi.org/10.1093/cercor/bhs206>
- PARKER, O. (1983). Quantitative differences in frequency perceptions by violinists, pianists, and trombonists. *Bulletin of the Council for Research in Music Education*, 76, 49–58.
- PATEL, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, 2, 142. <https://doi.org/10.3389/fpsyg.2011.00142>
- PATEL, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, 308, 98–108. <http://dx.doi.org/10.1016/j.heares.2013.08.011>
- RAVEN, J., RUST, J., CHAN, F., & ZHOU, X. (2018). *Raven’s 2 progressive matrices*. Pearson.
- ROGENMOSER, L., KERNBACH, J. M., SCHLAUG, G., & GASER, C. (2017). Keeping brains young with making music. *Brain Structure and Function*, 223, 297–305.
- SCELLENBERG, 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>
- 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., BOSNYAK, D. J., TRAINOR, L. J., & ROBERTS, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*, 23(13), 5545–5552. <https://doi.org/10.1523/jneurosci.23-13-05545.2003>

- SHI, J. (2016). *East meets West: A musical analysis of Chinese sights and sounds*, by Yuankai Bao. Louisiana State University and Agricultural & Mechanical College.
- SLATER, J., & KRAUS, N. (2015). The role of rhythm in perceiving speech in noise: A comparison of percussionists, vocalists and non-musicians. *Cognitive Processing*, 17, 79–87. <https://doi.org/10.1007/s10339-015-0740-7>
- STRANGE, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39(4), 456–466.
- TALAMINI, F., ALTOE, G., CARRETTI, B., & GRASSI, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLOS ONE*, 12(10), e0186773. <https://doi.org/10.1371/journal.pone.0186773>
- TERVANIEMI, M., JANHUNEN, L., KRUCK, S., PUTKINEN, V., & HUOTILAINEN, M. (2016). Auditory profiles of classical, jazz, and rock musicians: Genre-specific sensitivity to musical sound features. *Frontiers in Psychology*, 6, 1900. <https://doi.org/10.3389/fpsyg.2015.01900>
- TIERNEY, A., & KRAUS, N. (2014). Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience*, 8, 949. <https://doi.org/10.3389/fnhum.2014.00949>
- 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), 2657–2657. <https://doi.org/10.1038/s41598-023-29733-3>
- TONG, X., CHOI, W., & MAN, Y. Y. (2018). Tone language experience modulates the effect of long-term musical training on musical pitch perception. *Journal of the Acoustical Society of America*, 144(2), 690–697. <https://doi.org/10.1121/1.5049365>
- VUUST, P., BRATTICO, E., SEPPÄNEN, M., NÄÄTÄNEN, R., & TERVANIEMI, M. (2012). The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia*, 50(7), 1432–1443. <https://doi.org/10.1016/j.neuropsychologia.2012.02.028>
- WECHSLER, D. (2014). *Wechsler Adult Intelligence Scale* (4th ed., Hong Kong). Pearson Hong Kong.
- WESTFALL, P. J., JOHNSON, W. O., UTTS, J. M. (1997). A Bayesian perspective on the Bonferroni adjustment. *Biometrika*, 84(2), 419–427.
- WINSKEL, H. (2011). Orthographic and phonological parafoveal processing of consonants, vowels, and tones when reading Thai. *Applied Psycholinguistics*, 32(4), 739–759. <https://doi.org/10.1017/S014271641100004X>
- WONG, P. C. M., & PERRACHIONE, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28(4), 565–585. <https://doi.org/10.1017/S0142716407070312>
- ZHANG, J. D., SUSINO, M., MCPHERSON, G. E., & SCHUBERT, E. (2020). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music*, 48(3), 389–409. <https://doi.org/10.1177/0305735618804038>
- ZHENG, Y., & SAMUEL, A. G. (2018). The effects of ethnicity, musicianship, and tone language experience on pitch perception. *Quarterly Journal of Experimental Psychology*, 71(12), 2627–2642. <https://doi.org/10.1177/1747021818757435>
- ZUK, J., BENJAMIN, C., KENYON, A., & GAAB, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLOS ONE*, 9(6), e99868. <https://doi.org/10.1371/journal.pone.0099868>

Appendix A

Analytical Issues and Statistical Strategies

Our second research question was whether unpitched musicians performed similarly to nonmusicians in lexical tone discrimination and sequence recall. Statistically, null hypothesis significance testing either rejects or does not reject the null hypothesis. Thus, it cannot differentiate between absence of evidence and evidence of absence (Ly et al., 2019). Simply put, the (potential) lack of significant difference between the unpitched musicians and the nonmusicians (i.e., $p > .05$) would not necessarily indicate that the two groups performed similarly. In that case, we could only conclude that we had found no evidence that the two groups differed. To overcome this issue, we adopted Bayesian hypothesis testing (Dienes, 2014, 2016).

Bayesian hypothesis testing can provide both evidence of absence and presence. Unlike null hypothesis

significance testing, the Bayesian approach measures the strength of evidence for both the null and alternative hypotheses (Dienes, 2014, 2016). Bayes factor (BF_{10} or BF_{01}) is the ratio of the likelihood of the data given one hypothesis to the likelihood of the data given another hypothesis. For example, a BF_{10} of 30 indicates that the data are 30 times more likely under the alternative hypothesis than the null hypothesis. By contrast, a BF_{01} of 30 suggests that the data are 30 times more likely under the null hypothesis than the alternative hypothesis. Table A1 summarizes the Bayes factors and their respective evidence strength (Lee & Wagenmakers, 2013). Due to the lack of previous findings, we used a uniform prior that is relatively objective (Jaynes, 2003).

TABLE A1. Bayes Factors and Their Respective Evidence Strength

BF ₁₀	BF ₀₁	Hypothesis Supported	Evidence Strength
above 100	below 0.01	Alternative hypothesis	Extreme
30–100	0.01–0.033	Alternative hypothesis	Very strong
10–30	0.033–0.10	Alternative hypothesis	Strong
3–10	0.10–0.33	Alternative hypothesis	Moderate
1–3	0.33–1	Alternative hypothesis	Anecdotal
1	1	None	No evidence
0.33–1	1–3	Null hypothesis	Anecdotal
0.10–0.33	3–10	Null hypothesis	Moderate
0.033–0.10	10–30	Null hypothesis	Strong
0.01–0.033	30–100	Null hypothesis	Very strong
below 0.01	above 100	Null hypothesis	Extreme

Appendix B

Supplementary Analysis on the Interaction between Context and Group in Sequence Recall

To unpack the interaction between context and group, we conducted a Bayesian one-way ANCOVA in each group with mean accuracy as the dependent variable, context (vowel, level tone, and contour tone) as the within-subject factor, and working memory and non-verbal intelligence as the covariates. For the pitched musicians, the best-fit model was the null model. BF₀₁ indicated that the other models were 1.2–13.99 times less likely than the null model (see Table A6).

For the unpitched musicians, the best-fit model was the model (Context + IQ) with significant main effect of context and IQ. BF₀₁ revealed that the other models were 1.78–16.84 times less likely than the best-fit model (see Table A7). For the main effect of context, post hoc comparisons showed that the unpitched musicians performed significantly better in the vowel context than the level tone context and the contour tone context with

moderate evidence (BF_{10, U} = 4.49) and anecdotal evidence (BF_{10, U} = 1.39), respectively. They performed similarly across the level tone and contour tone context with anecdotal evidence, BF_{10, U} = 0.36 (see Table A8).

For the nonmusicians, the best-fit model was the model (Context) with a significant main effect of context. BF₀₁ revealed that the null model and the other models were 473.49 and 2.21–2111.09 times less likely than the best-fit model (see Table A9). For the main effect of context, post hoc comparisons indicated that the nonmusicians performed significantly better in the contour tone and vowel contexts than in the level tone context, with very strong (BF_{10, U} = 64.41) and extreme evidence (BF_{10, U} = 414.27), respectively. However, they performed similarly in the contour tone and vowel contexts with moderate evidence, BF_{10, U} = 0.25 (see Table A10).

TABLE A2. Model Comparison Relative to the Null Model of Discrimination Accuracy

Models	P(M)	P(M data)	BF _M	BF ₁₀	error %
Null model	0.050	1.103×10 ⁻³⁸	2.096×10 ⁻³⁷	1.000	
Tone + Group	0.050	0.444	15.195	4.027×10 ⁺³⁷	1.658
Tone	0.050	0.263	6.789	2.386×10 ⁺³⁷	0.368
Tone + Group + IQ	0.050	0.082	1.689	7.400×10 ⁺³⁶	10.239
Tone + Group + WM	0.050	0.077	1.576	6.943×10 ⁺³⁶	3.849
Tone + IQ	0.050	0.052	1.047	4.735×10 ⁺³⁶	0.949
Tone + WM	0.050	0.045	0.896	4.081×10 ⁺³⁶	1.430
Tone + Group + WM + IQ	0.050	0.020	0.394	1.839×10 ⁺³⁶	3.599
Tone + WM + IQ	0.050	0.014	0.274	1.290×10 ⁺³⁶	0.746
Tone + Group + Tone × Group	0.050	0.002	0.032	1.534×10 ⁺³⁵	3.725

Note. Showing the null model and the best nine other models, in descending order of BF₁₀. Tone = tone contrast; WM = working memory; IQ = nonverbal intelligence.

TABLE A3. Model Comparison Relative to the Null Model of Sequence Recall Accuracy

Models	P(M)	P(M data)	BF _M	BF ₁₀	error %
Null model	0.050	7.561×10^{-8}	1.437×10^{-6}	1.000	
Context + Group + IQ + Context×Group	0.050	0.421	13.828	$5.571 \times 10^{+6}$	2.471
Context + Group + Context×Group	0.050	0.300	8.161	$3.974 \times 10^{+6}$	1.300
Context + Group + WM + IQ + Context×Group	0.050	0.158	3.568	$2.091 \times 10^{+6}$	1.853
Context + Group + WM + Context×Group	0.050	0.092	1.919	$1.213 \times 10^{+6}$	1.988
Context + Group + IQ	0.050	0.012	0.227	156257.411	2.854
Context + Group	0.050	0.009	0.169	116441.408	1.091
Context + Group + WM + IQ	0.050	0.004	0.083	57712.880	1.006
Context + Group + WM	0.050	0.003	0.049	34211.242	1.794
Context + IQ	0.050	5.457×10^{-4}	0.010	7216.572	2.818

Note. Showing the null model and the best nine other models, in descending order of BF₁₀. Context = sequence recall context; WM = working memory; IQ = nonverbal intelligence.

TABLE A4. Model Comparison Relative to the Null Model of Vowel Sequence Recall Accuracy

Models	P(M)	P(M data)	BF _M	BF ₁₀	error %
Null model	0.125	0.020	0.142	1.000	
Group + IQ	0.125	0.338	3.569	17.020	1.015
Group	0.125	0.328	3.412	16.517	0.015
Group + WM + IQ	0.125	0.128	1.027	6.451	0.750
Group + WM	0.125	0.104	0.815	5.259	0.933
IQ	0.125	0.052	0.386	2.637	4.976×10^{-4}
WM + IQ	0.125	0.022	0.155	1.095	0.002
WM	0.125	0.008	0.060	0.427	0.002

Note. WM = working memory; IQ = nonverbal intelligence.

TABLE A5. Model Comparison Relative to the Null Model of Level Tone Sequence Recall Accuracy

Models	P(M)	P(M data)	BF _M	BF ₁₀	error %
Null model	0.125	1.252×10^{-4}	8.764×10^{-4}	1.000	
Group + IQ	0.125	0.396	4.590	3163.594	0.891
Group	0.125	0.367	4.052	2928.791	0.010
Group + WM + IQ	0.125	0.139	1.133	1112.509	0.890
Group + WM	0.125	0.097	0.752	775.049	1.380
IQ	0.125	6.838×10^{-4}	0.005	5.462	2.056×10^{-4}
WM + IQ	0.125	2.184×10^{-4}	0.002	1.745	9.871×10^{-4}
WM	0.125	3.794×10^{-5}	2.656×10^{-4}	0.303	0.003

Note. WM = working memory; IQ = nonverbal intelligence.

TABLE A6. Model Comparison Relative to the Best-fit Model of the Pitched Musicians

Models	P(M)	P(M data)	BF _M	BF ₀₁	error %
Null model	0.125	0.313	3.188	1.000	
IQ	0.125	0.261	2.470	1.200	2.506
WM + IQ	0.125	0.148	1.217	2.112	1.605
WM	0.125	0.145	1.192	2.151	1.493
Context	0.125	0.048	0.352	6.536	0.642
Context + IQ	0.125	0.039	0.287	7.939	2.015
Context + WM + IQ	0.125	0.023	0.165	13.612	2.498
Context + WM	0.125	0.022	0.160	13.991	2.389

Note. WM = working memory; IQ = nonverbal intelligence.

TABLE A7. Model Comparison Relative to the Best-fit Model of the Unpitched Musicians

Models	P(M)	P(M data)	BF _M	BF ₀₁	error %
Context + IQ	0.125	0.348	3.736	1.000	
Context	0.125	0.196	1.706	1.776	1.499
Context + WM + IQ	0.125	0.173	1.460	2.016	1.555
IQ	0.125	0.088	0.679	3.935	1.925
Context + WM	0.125	0.082	0.624	4.251	2.258
Null model	0.125	0.052	0.381	6.735	1.405
WM + IQ	0.125	0.041	0.298	8.534	1.486
WM	0.125	0.021	0.148	16.842	1.565

Note. WM = working memory; IQ = nonverbal intelligence.

TABLE A8. Post hoc Comparisons of the Unpitched Musicians' Accuracy Across Sequence Recall Contexts

		Prior Odds	Posterior Odds	BF _{10, U}	error %
Vowel	Level tone	0.587	2.640	4.494	4.130×10 ⁻⁷
	Contour tone	0.587	0.816	1.390	3.198×10 ⁻⁶
Level tone	Contour tone	0.587	0.210	0.358	0.018

TABLE A9. Model Comparison Relative to the Best-fit Model of the Nonmusicians

Models	P(M)	P(M data)	BF _M	BF ₀₁	error %
Context	0.125	0.461	5.992	1.000	
Context + WM	0.125	0.209	1.848	2.208	2.045
Context + IQ	0.125	0.204	1.799	2.256	1.402
Context + WM + IQ	0.125	0.123	0.986	3.735	2.390
Null model	0.125	9.740×10 ⁻⁴	0.007	473.488	0.806
IQ	0.125	3.931×10 ⁻⁴	0.003	1173.373	1.252
WM	0.125	3.920×10 ⁻⁴	0.003	1176.430	0.967
WM + IQ	0.125	2.185×10 ⁻⁴	0.002	2111.092	1.154

Note. WM = working memory; IQ = nonverbal intelligence.

TABLE A10. Post hoc Comparisons of the Nonmusicians' Accuracy Across Sequence Recall Contexts

		Prior Odds	Posterior Odds	BF _{10, U}	error %
Vowel	Level tone	0.587	243.343	414.271	5.894×10 ⁻⁸
	Contour tone	0.587	0.149	0.254	0.018
Level tone	Contour tone	0.587	37.833	64.408	2.284×10 ⁻⁷