

**Instrumental music training relates to intensity assessment but not emotional prosody
recognition in Mandarin**

Mengting Liu¹, Xiangbin Teng², Jun Jiang^{3*}

¹ Department of Art, Harbin Conservatory of Music, Harbin, China

² Department of Psychology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

³ Music College, Shanghai Normal University, Shanghai, China

* Corresponding author

E-mail: jjxytc@163.com (Jun Jiang)

Abstract

Building on research demonstrating the benefits of music training for emotional prosody recognition in nontonal languages, this study delves into its unexplored influence on tonal languages. In these languages, the acoustic similarity between lexical tones and music, and the dual role of pitch in conveying lexical and affective meanings, create a unique interplay. We evaluated 72 participants: half with extensive instrumental music training and the other half demographically matched controls. All participants completed an online test comprising 210 Chinese pseudosentences designed to express happiness, sadness, fear, anger, or neutrality. Our robust statistical analyses, incorporating effect size estimates and Bayesian factors, revealed that music and nonmusic groups exhibit similar abilities in identifying the emotional prosody of various emotions. However, compared to the nonmusic group, the music group attributed higher intensity ratings to emotional prosodies of happiness, fear, and anger. These findings suggest that while instrumental music training is not related to emotional prosody recognition, it does appear to be related to perceived emotional intensity. This dissociation between emotion recognition and intensity evaluation adds a new piece to the puzzle of the complex relationship between music training and emotion perception in tonal languages.

Keywords: speech, acoustic feature, emotion recognition

Introduction

Accurate recognition of emotional prosody is crucial for effective social interactions [1, 2]. Individuals adept at identifying others' emotional tones are better prepared for nuanced social engagement [3-6]. In contrast, those with challenges in understanding emotional prosody often face social complications [e.g., 7, 8-12]. Crucial is the examination of how individuals from varied linguistic backgrounds perceive emotional prosody. Exploring this, particularly the impact of specific training on social well-being through improved prosody recognition, unveils insights into cross-cultural variations in the utilization of acoustic cues, especially among individuals with and without tonal language experience [13, 14].

Recognizing emotional prosody hinges on acoustic features like pitch, intensity, tempo, and voice quality/timbre [15-17], which also serve as essential elements in music, such as Western tonal music. Different emotional prosodies are characterized by distinct features: for instance, happy prosody involves moderately high average pitch, high average voice intensity, and a rapid speech rate, whereas sad prosody is marked by low average pitch, low average voice intensity, and a slow speech rate [18-21]. Effective recognition of emotional prosody requires the efficient extraction of these acoustic markers and appropriate perceptual analysis. Given the shared acoustic elements between music and speech in conveying emotion [18, 22-24], instrumental music training could enhance speech perception [25], as posited by Patel's OPERA hypothesis [26, 27]. The OPERA hypothesis posits that shared sensory and cognitive mechanisms underlie both music and speech perception [28, 29], with music imposing more demanding conditions on these systems. Therefore, training in one domain (e.g., music listening) could have cross-modal benefits for the other (e.g., prosody recognition).

Furthermore, individuals with musical training and tonal languages listeners may exhibit advantages in perceiving lexical tones and musical pitch by paying greater attention to the relevant sound cues acquired in speech or music [13, 29]. Tonal languages listeners and musicians have an advantage in recognizing and learning lexical tones, exhibiting cross-domain transfer in pitch perception [13, 30]. Pitch plays a crucial role in both speech and music, serving as a predictive factor in the perception of emotional aspects of vocals. In nontonal languages, it is primarily utilized for emotional expression in speech. In contrast, tonal languages use pitch variations not only as emotional markers but also to convey semantic information [31-36]. As a result, the task of recognizing emotional prosody in tonal languages becomes more intricate, potentially requiring additional cognitive resources to mitigate interference from semantic aspects during emotion recognition. While music training is associated with enhanced perception of lexical tones in native speakers of tonal languages [37], it remains uncertain whether it is associated with improved perception of emotional prosody in their native language.

Many studies have examined the relationship between music training and emotional prosody recognition in native speakers of nontonal languages, such as English, German, Portuguese, Dutch, and Hebrew [38-41] (see [42-44], for reviews). Most of these studies utilize a cross-sectional design comparing musically trained and untrained individuals, while a few use a randomized experimental design, often randomly assigning musically untrained individuals to experiment and control groups (see S1 Table for an overview). Cross-sectional studies have yielded mixed findings on this relationship. Some studies demonstrate that adults with at least 6 years of instrumental music training [38-41, 45-48] and children with 3

years of extracurricular instruction [49] exhibit greater accuracy in identifying emotional prosodies. These benefits, however, do not extend to measures such as reaction times or emotional intensity ratings [39, 40, 50]. Contrarily, other research finds no significant benefit of music training in this domain, even with 5-14 years of instrumental practice [50-53]. These divergences may be attributable to varying sample sizes, impacting statistical significance, as well as the quantity of music training [54]. Specifically, studies which observed an effect of music training had larger sample sizes ($20 \leq n \leq 58$), making it easier to detect group differences. Meanwhile, studies which did not observe an effect of music training [50, 51, 53] had smaller sample sizes ($10 \leq n \leq 13$) and thus failed to obtain statistically significant results. This is because sample size is related to statistical significance (p -value); the smaller the sample size, the greater the p -value [55, 56]. However, the discrepancy between Mualem and Lavido's [52] study and other studies [45, 46] cannot be explained by sample size, and may be due to differences in the amount of music training. It is believed that at least 10 years of mixed music training is necessary to have an effect on auditory emotion recognition [54]. The average years of music training of the participants in the study of Mualem and Lavido was approximately 9 years, which may explain why their performance was not altered. Notably, the above research predominantly concentrates on nontonal languages, leaving a significant gap in understanding the implications for tonal languages like Mandarin, Thai, or Vietnamese.

Recognizing that lexical tones are a common element in both language and music [57], existing studies highlight how language and cultural backgrounds influence the processing of emotional cues. Notable differences exist in how listeners of tonal and nontonal languages

employ acoustic cues to interpret these tones [58, 59]. Furthermore, East Asians are more sensitive to verbal cues of an emotional utterance, whereas Westerners rely more on semantic cues for understanding and interpreting emotional information [60-62]. While longitudinal experimental studies have demonstrated that musical training enhances the recognition of emotional prosody in children [46, 63] and young adults [52, 64] but not in older adults [65, 66] who speak nontonal languages, the link between musical training and the perception of emotional prosody in speakers of tonal languages remains insufficiently established. To fill this gap in current research, the present study aims to explore the connection between music training and the recognition of emotional prosody in tonal languages. For this purpose, undergraduate students majoring in music performance or musicology, all of whom had received instrumental training since childhood, were compared with students without professional instrumental training. The focus was on their ability to recognize emotional prosody—specifically happiness, sadness, fear, anger, and neutrality—in Mandarin pseudosentences. Building on insights from existing studies [38-41], we hypothesized that music majors would excel in accurately identifying Mandarin emotional prosody compared to nonmusic majors.

Method

Participants

This study was approved by the Ethics Committee of the Harbin Conservatory of Music on 04/06/2023 and all participants provided verbal informed consent form and received monetary compensation for their participation. To achieve 90% power to detect a large main effect ($d = 0.8$) of group (music vs. nonmusic) at the 0.05 level, a minimum sample size of 34

for each group was required using G*Power [Version 3.1.9.7; 67]. Consequently, 36 music students and 36 nonmusic students who were native Mandarin speakers were recruited for the study from 04/08/2023 to 05/29/2023. The music students had been playing their chosen instruments for a minimum of 12 years, with an average of 13.9 hr of practice per week, or at least 3 hr of practice per day [54, 68]. The instruments covered a wide range, including piano ($n = 22$), violin ($n = 4$), double bass ($n = 1$), viola ($n = 1$), flute ($n = 1$), clarinet ($n = 1$), oboe ($n = 1$), harp ($n = 1$), yangqin ($n = 1$), trumpet ($n = 1$), suona ($n = 1$), and erhu ($n = 1$). In contrast, the nonmusic students had only received compulsory music education in school. To ensure differences in self-reported musical skills and behaviours between music and nonmusic students, the Goldsmiths Musical Sophistication Index [Gold-MSI; 69] was administered. As shown in Table 1, there was moderate evidence that the two groups were matched for age, gender, years of education, nonverbal IQ as assessed by the Matrix Reasoning Item Bank [MaRs-IB; 70]; and handedness as assessed by the Edinburgh Handedness Inventory [71], but very strong evidence that the musician group scored higher than the nonmusic group on all Gold-MSI subtests and the general musical sophistication factor. All participants had normal hearing and no history of neurological or psychiatric disorders.

Table 1. Participant characteristics

Variables	Music	Nonmusic	Statistics	Effect size	95% CI	<i>BF</i>
Age (years)	22.08 ± 2.17	22.31 ± 1.95	$t(69.25) = -0.46$	$g^* = -0.11$	[-0.56, 0.35]	0.13
Sex (female/male)	34/2	34/2	Fisher's exact test, $p = 1.000$	$\log OR = 0$	[-2.68, 2.68]	0.14
Education (years)	15.94 ± 2.34	16.11 ± 2.04	$t(68.67) = -0.32$	$g^* = -0.08$	[-0.53, 0.38]	0.12
MaRs-IB (%)	77.53 ± 10.95	77.42 ± 11.45	$t(69.86) = 0.04$	$g^* = 0.01$	[-0.45, 0.47]	0.12
Handedness	81.02 ± 16.89	83.27 ± 15.19	$t(69.22) = -0.59$	$g^* = -0.14$	[-0.60, 0.32]	0.14
Years of music training	15.42 ± 2.25	—	$t(35) = 41.16$	$g = 6.71$	[5.11, 8.27]	6.27×10^{27}
Gold-MSI						
Musical Training	5.56 ± 0.47	1.31 ± 0.25	$t(53.69) = 48.22$	$g^* = 11.21$	[8.99, 13.32]	∞
Perceptual Abilities	5.98 ± 0.67	3.73 ± 0.95	$t(62.95) = 11.65$	$g^* = 2.71$	[2.05, 3.37]	3.61×10^{28}
Singing Abilities	5.33 ± 0.74	2.67 ± 1.07	$t(62.36) = 12.29$	$g^* = 2.86$	[2.18, 3.53]	7.50×10^{31}
Active Engagement	5.24 ± 0.68	3.10 ± 1.02	$t(60.88) = 10.54$	$g^* = 2.45$	[1.82, 3.08]	1.55×10^{23}
Emotions	5.76 ± 0.73	4.35 ± 0.97	$t(64.84) = 6.97$	$g^* = 1.62$	[1.08, 2.15]	4.07×10^9
General Sophistication	5.40 ± 0.51	2.58 ± 0.77	$t(60.60) = 18.38$	$g^* = 4.28$	[3.39, 5.16]	2.65×10^{72}

Stimuli

This study utilized emotional prosody materials drawn from the Chinese vocal emotional stimuli database [72]. The database features 874 pseudosentences, produced by four native Mandarin-Chinese speakers (two male), each conveying one of seven emotions (happiness, sadness, fear, anger, disgust, pleasant surprise, and neutrality) through prosodic variation in 35 unique utterances. Each utterance, ranging from 1 to 2 s in duration, adhered to Chinese grammatical rules while remaining semantically meaningless to eliminate the influence of semantics on emotional prosody perception [72-74]. The stimuli were validated with a group of native Mandarin speakers who participated in a seven-option forced choice task to identify the expressed emotion and a 5-point Likert item (1 = *very weak*, 5 = *very intense*) to rate its intensity, excluding neutrality. Stimuli meeting satisfactory identification accuracy criteria were included in the database; specifically, those exceeding a 42.86% average recognition rate (3 times the chance level) for the target emotion and remaining below 42.86 % for any nontarget emotion [72, 75, 76].

Twenty-one utterances each expressing happiness, sadness, fear, anger and neutrality were selected from recordings of a male and a female speaker, yielded 42 stimuli per emotion. Table 2 presents the recognition rate and intensity rating for each emotion except neutrality. A one-way analysis of variance (ANOVA) was conducted on the average accuracy rates. Although normality assumptions were met ($W_s \geq 0.94$, $ps \geq .200$), sphericity was violated ($W = 0.26$, $p = .003$). Therefore, the degrees of freedom were corrected using the Greenhouse-Geisser method ($\varepsilon = 0.709$). The analysis revealed moderate evidence for a main effect of emotion with a large effect size, $F(2.83, 56.70) = 5.87$, $\omega_p^2 = 0.185$, 95% CI [0.018,

0.343], $BF = 6.09$. Pairwise comparisons (see S2 Table) showed that happiness was recognized less accurately than sadness, anger, and neutrality. No differences were found between any other emotion pairs. Similar analyses were conducted for the average intensity ratings. The data met normality assumptions ($W_s \geq 0.94$, $ps \geq .223$) but violated sphericity ($W = 0.55$, $p = .048$). ANOVA with the Greenhouse-Geisser correction ($\varepsilon = 0.703$) revealed very strong evidence for a main effect of emotion with a large effect size, $F(2.11, 42.20) = 14.50$, $\omega_p^2 = 0.386$, 95% CI [0.150, 0.557], $BF = 1.52 \times 10^3$. Pairwise comparisons (see S2 Table) indicated that happiness was rated as less intense than sadness, fear, and anger, while sadness was rated as more intense than fear. No differences were found between anger and sadness or fear.

Table 2. Acoustic and perceptual measures for each emotion category averaged across two speakers ($n = 21$)

Measure	Happiness		Sadness		Fear		Anger		Neutrality	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Acoustic features										
f_0 mean (Hz)	1.58	0.13	0.78	0.07	1.14	0.13	1.54	0.15	0.40	0.06
f_0 range (Hz)	2.25	0.32	0.84	0.31	1.22	0.26	2.05	0.32	0.93	0.24
Amp mean (dB)	0.57	0.04	0.63	0.05	0.58	0.05	0.55	0.04	0.56	0.05
HNR mean (dB)	12.16	1.43	14.55	1.23	11.92	1.77	9.22	1.56	10.15	1.48
HNR <i>SD</i> (dB)	5.81	0.65	6.84	0.54	6.87	0.73	5.55	0.56	6.52	0.71
Speech rate (syllables/s)	5.99	0.65	5.12	0.40	5.88	0.41	7.01	0.66	6.23	0.47
Perceptual features										
Recognition rate	0.72	0.10	0.83	0.05	0.77	0.11	0.81	0.07	0.83	0.09
Intensity rating	2.64	0.33	3.15	0.26	2.94	0.28	3.01	0.19	—	—

Note. The f_0 and Amp values in the original data set were normalized.

To assess whether the main acoustic features (mean fundamental frequency [f_0 mean], f_0 range, mean amplitude [Amp mean], mean harmonics-to-noise ratio [HNR mean], HNR variation [HNR *SD*], and speech rate) of the stimuli (see Table 2) could predict the intended emotion categories, we conducted a discriminant analysis following the procedures of Liu and Pell [72] and Paulmann and Uskul [77]. Using IBM SPSS Statistics (Version 29.0.2), we treated the main acoustic features as predictor variables and the intended emotions as the outcome variable for the pseudosentences. The analysis revealed that 84.76% (178/210) of the pseudosentences were correctly classified in terms of their intended emotion: happiness 78.57% (33/42), sadness 97.62% (41/42), fear 76.19% (32/42), anger 80.95% (34/42), and neutrality 90.48% (38/42). These results suggest that listeners can reliably identify the intended emotions based on the analyzed acoustic features.

Measures

MaRs-IB

The MaRs-IB was employed to measure participants' nonverbal IQ [70]. This open-access online assessment has demonstrated good reliability and validity, and is similar to Raven's Advanced Progressive Matrices [78]. The test consists of 80 items, each featuring a 3×3 matrix in which eight out of nine cells contain abstract shapes that vary along four dimensions: color, size, shape and position. Participants must select the missing shape from an array of four options. Each item answered correctly is worth one point, while incorrect items are worth zero points. A higher score or percentage of accuracy is indicative of a stronger spatial reasoning ability and a higher nonverbal IQ [79].

Gold-MSI

The Gold-MSI [69] is a questionnaire designed to assess individual differences in self-reported musical skills and behaviors among both musicians and the general population. It consists of 38 items divided into five subscales: Musical Training (7 items), Perceptual Abilities (9 items), Singing Abilities (7 items), Active Engagement (9 items), and Emotions (6 items). Respondents are asked to rate their agreement with each statement on a 7-point Likert scale (from 1 *strongly disagree* to 7 *strongly agree*) for the first 31 items or select one of seven options for the remaining 7 items. The scores for each dimension are then summed, with higher scores indicating a higher level of musical skills and greater musical engagement or emotional experiences. Additionally, the General Sophistication consists of 18 items, with higher scores indicating a greater degree of musical sophistication. The Chinese version of the Gold-MSI has been found to demonstrate high reliability and validity [80, 81] and was used in the present study. In the current study, the reliability of the Musical Training, Perceptual Abilities, Singing Abilities, Active Engagement, Emotions, and General Sophistication factors, as evaluated by McDonald's ω , was 0.98 (95% CI [0.97, 0.99]), 0.95 (95% CI [0.93, 0.96]), 0.94 (95% CI [0.91, 0.96]), 0.93 (95% CI [0.91, 0.96]), 0.84 (95% CI [0.78, 0.90]), and 0.97 (95% CI [0.96, 0.98]) respectively.

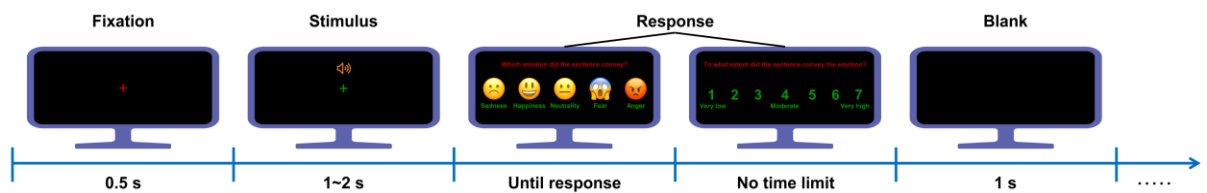
Procedure

Participants completed the online experiment through the NAODAO platform (www.naodao.com) between 04/15/2023 and 06/06/2023. Before the experiment began, participants were required to read and fully understand the informed consent form. The experiment only commenced automatically upon their explicit confirmation of participation.

To ensure participants were familiar with the procedure, a practice session preceded the formal experiment. This practice session included four practice stimuli and feedback on accuracy. Additionally, participants could adjust the headphone volume to a comfortable level.

Fig 1 displays the design of the trial in the experiment. At the start of each trial, a red fixation cross appeared in the center of the screen for 0.5 s to indicate that a stimulus was about to come. The fixation cross then turned green as the stimulus began to play for approximately 1-2 s. After listening to the stimulus, a response interface appeared and participants were asked to click on one of the five emotional prosody labels and icons to match the emotion expressed in the speech as fast as possible without sacrificing accuracy. Subsequently, a 7-point Likert item of emotion intensity (1 = *very low*, 7 = *very high*) was presented on the screen, and participants were asked to rate the intensity of the emotion in the stimulus without a time limit. To control for potential gender-related differences in vocal acoustic features [82, 83] and reduce participants' cognitive load from sudden change in speaker gender, we presented sentences spoken by male and female speakers in two separate blocks, with 105 randomly presented sentences per block. The order of the blocks was counterbalanced across participants, with a 1-2 min break between the two blocks.

Fig 1. Timing of one trial used in the experiment



Data analysis

Before data analysis, trials with response times (RTs) falling below 0.2 s (anticipatory

responses or fast guesses) or exceeding 2 *SD* above the mean for each participant (distraction) were excluded as outliers. Trials with errors were also omitted for RT and intensity ratings. However, in speeded decision tasks, a pervasive speed-accuracy tradeoff [SAT; 84, 85] exists. Analyzing mean RT or accuracy rates separately could, therefore, lead to contradictory or invalid conclusions [86-88]. To control for this potential SAT, a composite metric incorporating speed and accuracy should be used for performance evaluation. Composite metrics offer more sensitive and reliable assessments than relying solely on accuracy or RT measures [89, 90].

Several such metrics exist, including inverse efficiency score, rate correct score, linear integrated speed-accuracy score, and balanced integration score (BIS). Among these, only the BIS exhibits relatively insensitivity to SAT [91, 92]. To minimize the SAT effect and preserve “true” effects, we opted for the BIS, assigning equal weights to speed and accuracy. The BIS is calculated by subtracting the standardized (i.e., *z* scores) mean correct RT from the standardized mean proportions of correct responses across all conditions and participants [91-93]. Previous studies have widely adopted this metric [e.g., 94, 95-102]. Using this approach, participants achieving high accuracy at the expense of RT or fast RTs at the expense of accuracy will have BISs near zero compared to the sample average. Participants with both high accuracy and fast RTs will have higher positive BISs, while those with low accuracy and slow RTs will have lower negative BISs. Therefore, higher BIS represents faster and more accurate performance. To investigate the SAT in this study, we first analyzed accuracy and RT data. If an SAT is identified, we will subsequently analyze the data using the BIS framework. Both the frequentist and Bayesian approaches were applied to the data.

Frequentist analysis

In frequentist analysis, effect sizes and confidence intervals are used to evaluate the magnitude and precision of estimation of an effect and to draw conclusions, instead of relying on p -values from traditional null hypothesis significance tests [103-108]. This method of interpretation is endorsed by the American Psychological Association [109, 110] and several journals [111-113] and researchers [114-117]. If the confidence interval for an effect size does not include zero, the effect is considered to be present; while if the confidence interval (CI) does include zero, the effect is considered to be nonexistent [118-121]. Additionally, effect sizes can be categorized as small, medium, or large based on established thresholds: g or $d = 0.20, 0.50, \text{ and } 0.80$ [122], $r = 0.10, 0.30, \text{ and } 0.50$ [123], and $\omega^2 = 0.01, 0.06, \text{ and } 0.14$ [124]. This study employed effect sizes and their confidence intervals to interpret the results.

We analyzed the demographic variables of age, years of education, MaRs-IB, handedness, and Gold-MSI using Welch's t -tests with JASP (Version 0.18.3). Hedges' g^* and their 95% CIs [125] were calculated using the *deffectsiz* package (Version 0.0.0.9000) in R (Version 4.4.1) and RStudio (Version 2024.04.2+764). For gender, we used Fisher's exact test in JASP and calculated log OR and its 95% CI. For years of music training, we performed a one-sample t -test using the *statsExpressions* package (Version 1.5.4) and calculated g and its 95% CI.

For each perceptual feature, we conducted an ANOVA with emotion (happiness, sadness, fear, anger, and/or neutrality) as a within-items variable using jamovi (Version 2.5.6). Prior to the ANOVAs, we tested for normality with the Shapiro-Wilk test and sphericity with Mauchly's test. If assumptions were met, parametric tests were conducted. If

any assumption was violated, the Greenhouse-Geisser correction (ϵ) was applied to adjust the degrees of freedom.

BIS was used to assess overall task performance, but we also conducted a correlation analysis between accuracy and RT to test the SAT. We used repeated measures correlation [126-128] implemented in the *rmcorr* package (Version 0.6.0). The effect size (correlation coefficient) and its 95% CI were estimated using the bootstrapping with the suggested number of 10000 bootstrap resamples [129-132].

For accuracy, RT, BIS and emotional intensity, we conducted a two-way mixed design ANOVA with group (music and nonmusic) as the between-subjects variable and emotion (happiness, sadness, fear, anger, and neutrality) as the within-subjects variable. Data were tested for normality, homogeneity of variance with Levene's test, and sphericity using JASP. Parametric tests were conducted if assumptions were met; otherwise, nonparametric tests were conducted with *ARTool* (Version 0.11.1). Post hoc comparisons with Bonferroni correction were performed using the ART-C program for significant interaction effects. We used the *effectsize* package (Version 0.8.8) to calculate ω_p^2 and its 95% CI for each ANOVA effect, and d and its 95% CI for each pairwise or post hoc comparison.

Bayesian analysis

To assess the strength of the evidence and obtain more conservative results, we conducted Bayesian hypothesis testing. The Bayes factor (BF) was calculated to quantify the relative plausibility of the null hypothesis (H_0) and alternative hypothesis (H_1) given the observed data [133-136]. Conventionally, a BF of 1-3 indicates weak evidence for H_1 , 3-10 indicates moderate evidence, 10-30 indicates strong evidence, and above 30 indicates very

strong evidence for H_1 . Conversely, BF s between 0.33-1, 0.10-0.33, 0.03-0.10, and below 0.03 indicate weak, moderate, strong, and very strong evidence for H_0 , respectively [137, 138]. While acknowledging the inherent subjectivity in interpreting evidential strength, this study adopted the common practice of considering moderate evidence as a sufficient threshold for drawing meaningful conclusions. For example, a BF of 5 indicates that the data are 5 times more likely under H_1 than H_0 , implying substantial support for the presence of an effect.

For age, years of education, MaRs-IB, handedness and Gold-MSI, we employed the Bain module in JASP to perform Bayesian Welch's t -tests [139] and calculate BF s. For gender, we employed JASP to perform a Bayesian contingency table test based on an independent multinomial sampling scheme [140] and obtain the BF . For years of music training, we used JASP to execute a Bayesian one-sample t -test with the default prior (Cauchy $r = 0.707$) and determine the BF .

For perceptual features, accuracy, RT, BIS, or emotional intensity, we calculated BF s using *anovaBFcalc* (version 0.1.0) with default $\alpha = -0.5$ (MWS method) based on F -values and degrees of freedom obtained from frequentist analysis. For pairwise or post hoc comparisons, we computed BF s using the *BayesFactor* package (version 0.9.12-4.7) with a prior scale factor of $r = 0.707$, based on t values and sample sizes obtained from frequentist analysis.

Results

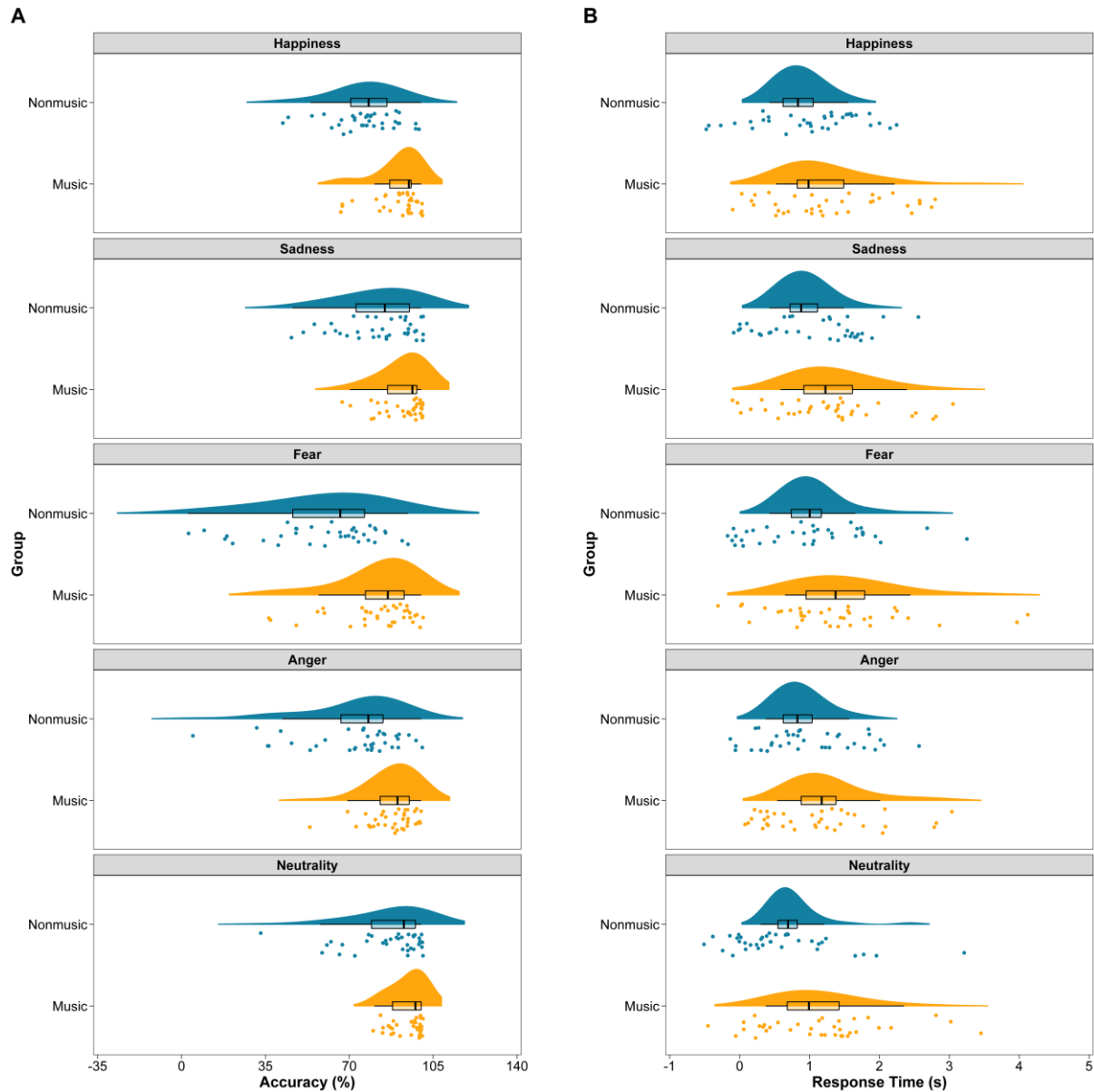
We removed 24 music group trials (0.16%) and 86 nonmusic group trials (0.57%) with RT < 200 ms. Subsequently, we discarded 312 music group trials (2.08%) and 368 nonmusic group

trials (2.45%) with $RT > (M + 2 SD)$ ms. Finally, we eliminated 753 music group trials (5.25%) and 1742 nonmusic group trials (12.16%) due to error responses.

Accuracy and response time

Regarding the accuracy data, all dataset except for happiness in the nonmusic group ($W = 0.95, p = .146$) deviated from normality ($Ws \leq 0.93, ps \leq .025$). Additionally, all datasets except for happiness ($F(1, 70) = 2.40, p = .126$) violated the assumptions of homogeneity of variance ($Fs(1, 70) \geq 5.01, ps \leq .028$) and sphericity ($W = 0.54, p < .001$). As for the RT data, all datasets except for happiness ($W = 0.95, p = .138$) and sadness ($W = 0.95, p = .121$) in the nonmusic group violated the normality assumption ($Ws \leq 0.93, ps \leq .024$). Furthermore, all RT datasets violated the assumptions of homogeneity of variance ($Fs(1, 70) \geq 4.47, ps \leq .038$) and sphericity ($W = 0.65, p < .001$).

Fig 2 illustrates the accuracy and RTs of the music and nonmusic groups across all emotions. It is evident that the music group exhibited higher accuracy but slower RTs compared to the nonmusic group for all emotions. To validate these visual observations, separate two-way ANOVAs were conducted on the accuracy and RT data. There was very strong evidence for a large main effect of group on accuracy ($F(1,70) = 62.10, \omega_p^2 = 0.459$, 95% CI = [0.290, 0.589], $BF = 3.13 \times 10^8$), indicating that the music group ($M = 89.47, SD = 11.67$) performed better than the nonmusic group ($M = 75.39, SD = 20.34$). Similarly, there was very strong evidence for a large main effect of group on RTs ($F(1,70) = 14.33, \omega_p^2 = 0.156$, 95% CI = [0.033, 0.312], $BF = 59.08$), with the music group exhibiting slower RTs ($M = 1.28, SD = 0.60$) than the nonmusic group ($M = 0.89, SD = 0.37$). These results suggest the presence of an SAT for both music and nonmusic subjects.

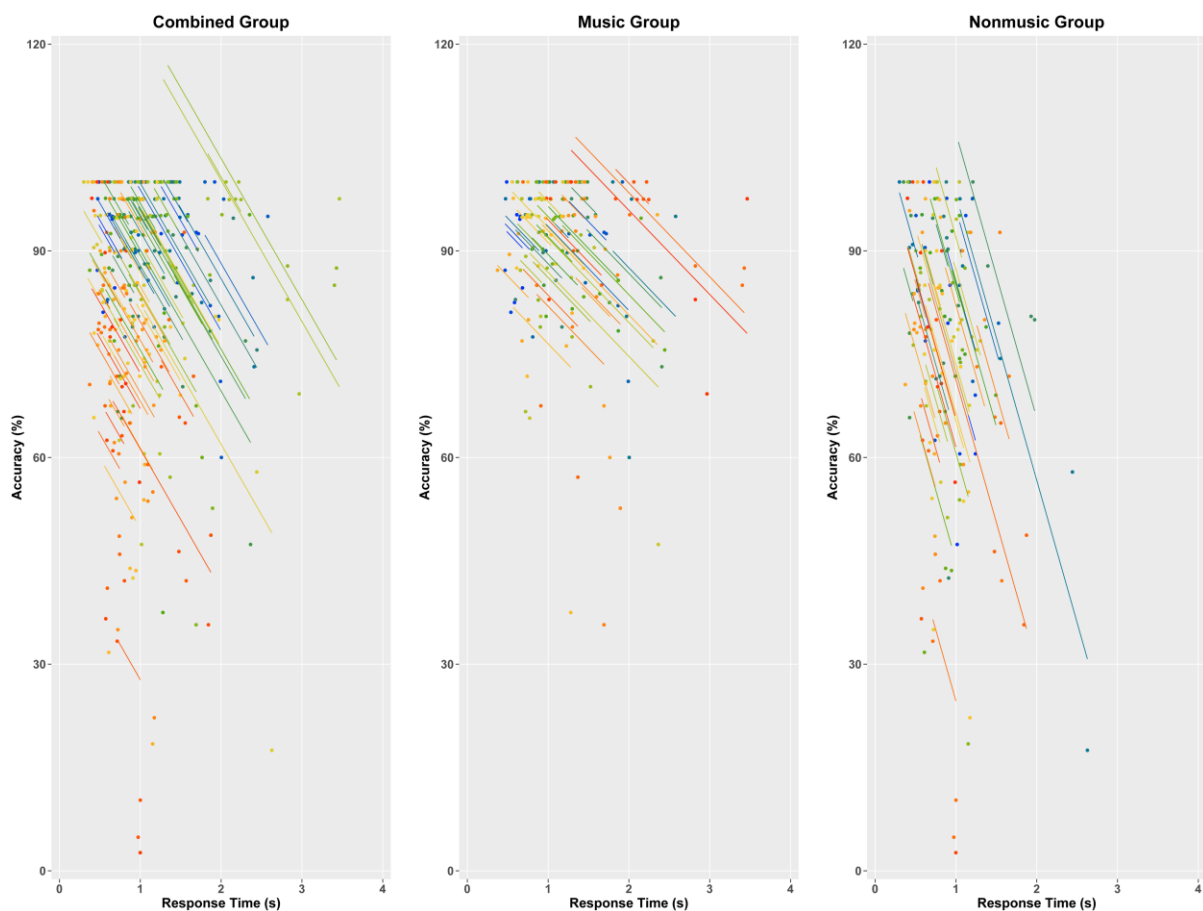
Fig 2. Accuracy (A) and response time (B) as a function of group and emotion

Note. These raincloud plots, combining a halved violin plot, a boxplot, and jittered individual data points, depict the data distribution. They were generated using *ggplot2* (Version 3.5.1), *PupillometryR* (Version 0.0.5), *ggpp* (Version 0.5.8-1), and *cowplot* (Version 1.1.3) in R.

Fig 3 depicts the speed–accuracy correlations for the combined group and each individual group across all emotions. A moderate negative correlation between RT and accuracy was observed for the combined group, $r(287) = -0.397$, 95% CI $[-0.485, -0.306]$, consistent with the SAT phenomenon. This correlation remained similar for the music group,

$r(143) = -0.366$, 95% CI $[-0.504, -0.204]$, but was larger for the nonmusic group, $r(143) = -0.510$, 95% CI $[-0.616, -0.392]$. These findings suggest that faster RT generally come at the expense of reduced accuracy within individuals, highlighting the suitability of BIS as a more comprehensive measure of task performance.

Fig 3. Scatterplots depicting correlations between response time and accuracy for the combined group and each group across emotion



Note. Each dot represents the average response time and accuracy for a specific emotion.

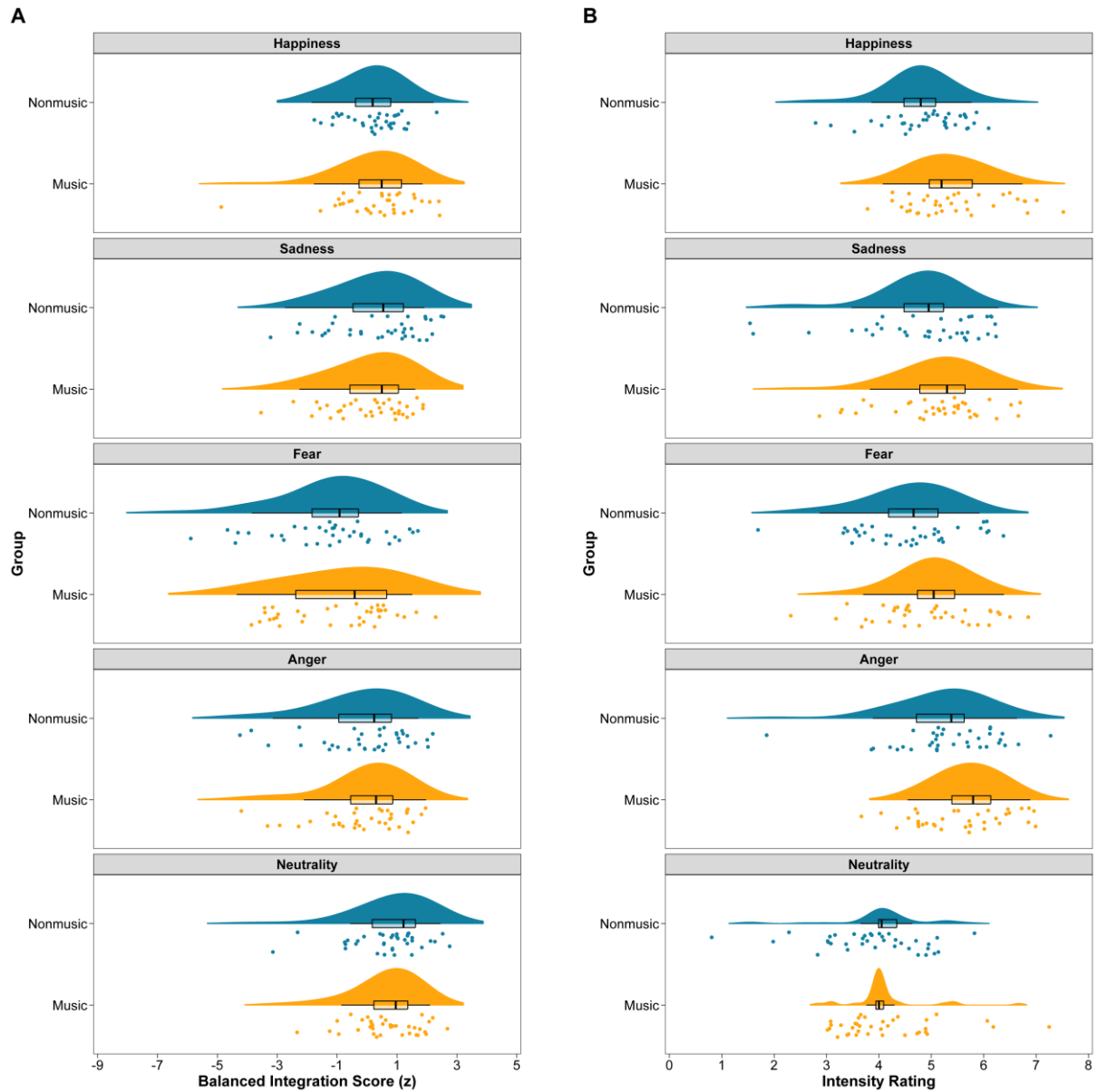
Participant identity is indicated by color, and colored lines show the repeated measures correlation fits for each participant's data. These scatterplots were created using the following R packages: *rmcorr* (Version 0.6.0), *ggplot2* (Version 3.5.1), *pals* (Version 1.8), and *cowplot* (Version 1.1.3).

Balanced integration score

All but one dataset (happiness for the nonmusic group: $W = 0.96, p = .281$) deviated from normality ($W_s \leq 0.94, p_s \leq .037$). Homogeneity of variance was confirmed for all datasets ($F_s(1, 70) \leq 0.76, p_s \geq .388$), but the data violated the assumption of sphericity ($W = 0.62, p < .001$).

Fig 4A illustrates the BIS of the music and nonmusic groups across all emotions, suggesting comparable performance. A nonparametric ANOVA corroborated these observations. There was substantial evidence for the absence of a main effect of group, $F(1, 70) = 0.14, \omega_p^2 = 0.000, 95\% \text{ CI} = [0.000, 0.000], BF = 0.103$, indicating that the music group ($M = 0.03, SD = 1.40$) did not outperform the nonmusic group ($M = -0.03, SD = 1.19$). In contrast, there was very strong evidence for a large main effect of emotion, $F(4, 280) = 34.62, \omega_p^2 = 0.321, 95\% \text{ CI} = [0.230, 0.396], BF = 8.41 \times 10^{19}$. Pairwise comparisons (see S3 Table) showed that participants performed best in detecting neutral emotional prosody ($M = 0.74, SD = 1.19$), followed by sadness ($M = 0.19, SD = 1.20$), happiness ($M = 0.17, SD = 1.08$), or anger ($M = -0.08, SD = 1.40$), and worst in detecting fear ($M = -1.03, SD = 1.70$). Notably, there was very strong evidence for the absence of an interaction effect between group and emotion, $F(4, 280) = 2.07, \omega_p^2 = 0.015, 95\% \text{ CI} = [0.000, 0.041], BF = 2.21 \times 10^{-3}$. These findings suggest that instrumental training is not associated with the ability to recognize emotional prosody in tonal languages.

Fig 4. Balanced integration score (A) and intensity ratings (B) as a function of group and emotion



Note. These raincloud plots, combining a halved violin plot, a boxplot, and jittered individual data points, depict the data distribution. They were generated using *ggplot2* (Version 3.5.1), *PupillometryR* (Version 0.0.5), *ggpp* (Version 0.5.8-1), and *cowplot* (Version 1.1.3) in R.

Intensity ratings

All but four datasets (neutrality for the music group, and sadness, anger, and neutrality for the nonmusic group: $W_s \leq 0.90$, $p_s \leq .004$) exhibited normality ($W_s \geq 0.94$, $p_s \geq .060$). Homogeneity of variance was met for all datasets ($F_s(1, 70) \leq 3.02$, $p_s \geq .087$), but the assumption of sphericity was violated ($W = 0.35$, $p < .001$).

Fig 4B demonstrates the intensity ratings of the music and nonmusic groups across all emotions, suggesting higher ratings for the music group. A nonparametric ANOVA revealed moderate evidence for a medium main effect of group, $F(1,70) = 8.47$, $\omega_p^2 = 0.094$, 95% CI = [0.006, 0.240], $BF = 4.92$. This indicates that the music group ($M = 5.07$, $SD = 0.85$) assigned higher intensity ratings than the nonmusic group ($M = 4.68$, $SD = 0.89$). Additionally, a very strong evidence for a large main effect of emotion emerged, $F(4, 280) = 74.55$, $\omega_p^2 = 0.508$, 95% CI = [0.428, 0.608], $BF = 3.24 \times 10^{39}$. Pairwise comparisons (see S3 Table) showed that angry emotional prosody received the highest rating ($M = 5.47$, $SD = 0.79$), followed by happy ($M = 5.05$, $SD = 0.71$) and sad ($M = 4.97$, $SD = 0.84$) or fearful ($M = 4.81$, $SD = 0.75$), with neutral receiving the lowest rating ($M = 4.09$, $SD = 0.77$). Furthermore, moderate evidence was found for a medium interaction effect between group and emotion, $F(4, 280) = 6.15$, $\omega_p^2 = 0.067$, 95% CI = [0.014, 0.121], $BF = 4.80$. Post hoc comparisons (Table 3) revealed very strong evidence for higher ratings by the music group for happy emotion, moderate evidence for higher ratings for fearful and angry emotions, similar ratings for neutral emotion, and only weak evidence for higher ratings for sad emotion. These findings suggest that instrumental training is associated with heightened sensitivity to the intensity of emotional prosodies in tonal languages.

Table 3. Post hoc comparisons for the Group \times Emotion interaction on intensity ratings

Emotion	Music	Nonmusic	$t(189.57)$	d	95% CI	BF
Happiness	5.32 ± 0.62	4.77 ± 0.69	3.51	0.51	[0.22, 0.80]	37.43
Sadness	5.16 ± 0.82	4.77 ± 0.82	2.37	0.34	[0.06, 0.63]	2.58
Fear	5.03 ± 0.67	4.59 ± 0.78	2.57	0.37	[0.09, 0.66]	3.87
Anger	5.72 ± 0.56	5.21 ± 0.90	2.95	0.43	[0.14, 0.72]	9.03

Neutrality	4.11 ± 0.67	4.06 ± 0.86	-0.58	-0.08	[-0.37, 0.20]	0.28
------------	-----------------	-----------------	-------	-------	---------------	------

Note. Only the $M \pm SD$ was computed from the raw data; others statistics were derived from the aligned-and-ranked data.

Discussion

This study uniquely explores the link between instrumental music training and emotional prosody identification in tonal languages, focusing on Mandarin. Utilizing a robust analysis with effect size estimates and Bayesian factors, we draw strong conclusions from our sizable sample of 72 participants. Interestingly, the results revealed no difference in emotional prosody identification between musically trained and untrained participants. However, music majors assigned higher intensity ratings to happy, fearful, and angry prosodies. These findings suggest that music training is not positively correlated with an improved ability to recognize emotional prosody in tonal languages, but is instead associated with perceived emotional intensity. This adds a new dimension to our understanding of the intricate relationship between music training and emotion perception in tonal languages, distinct from research focused on nontonal languages.

Utilizing the BIS to mitigate speed-accuracy tradeoff effects in our data, we discovered that music and nonmusic majors exhibited similar proficiency in recognizing emotional prosodies. This outcome contrasts with the studies suggesting that instrumentalists [38-41, 48] and mixed musicians [45, 46, 49] exhibit superior performance on emotional prosody recognition in nontonal languages. This inconsistency may be explained by the intricate interplay between music and language experience in shaping emotional prosody recognition. Indeed, prior research has demonstrated that musicians with a nontonal language background

(i.e., English) outperform their nonmusician counterparts in tasks involving linguistic pitch perception [141, 142] and lexical tone perception [14, 143-145]. However, this advantage is not as pronounced among musicians with a tonal language background such as Mandarin [146], Cantonese [147], and Thai [144, 148]. The null result observed in this study may be related to the characteristics of the music subjects. Specifically, the music subjects in our study are more likely to possess absolute pitch—the ability to name a tone without reference—due to their experience with a tonal language. Although we did not directly assess absolute pitch ability, previous research indicates that 53%-72% of Chinese music students possess absolute pitch [149-152]. Given that emotional prosody perception primarily relies on relative pitch processing [153, 154], which focuses on the pitch changes and patterns rather than absolute pitch, the presence of absolute pitch ability may not necessarily confer an advantage in emotional prosody recognition tasks. In contrast, among musical subjects who spoke a nontonal language, only 3%-10% have absolute pitch [149, 151, 155, 156], while the majority possess relative pitch. This heightened capacity for relative pitch processing [157, 158] likely contributes to their enhanced recognition of emotional prosody.

We also discovered that music majors gave higher intensity ratings to happy, fearful, and angry emotional prosody compared to nonmusic majors. This observation aligns with previous evidence that musicians provided higher intensity ratings for happy emotional prosody in English than did nonmusicians [159]. However, it is contradictory to the studies reporting no differences in intensity or arousal ratings for emotional prosody in Portuguese or German between musicians and nonmusicians [40, 160]. The discrepancy may be attributed to the difference in sample sizes; our study included 36 music and 36 nonmusic subjects,

whereas Dibben et al.'s and Lima and Castro's studies involved 17 music and 15 nonmusic subjects, and 20 music and 20 nonmusic subjects, respectively. The larger sample size in our study, being at least 1.8 times larger than those of the other studies, likely facilitated a more sensitive detection the effect of music training. Our findings might be explained by enhanced auditory processing and selective attention abilities in music subjects. Through extensive instrument training and practice, music subjects develop advanced auditory skills that make them more sensitive to subtle changes in acoustic features (e.g., pitch, intensity, and tempo) shared by speech and music. Studies show musicians excel in pitch [161-165], intensity [166], and tempo [167, 168] discrimination compared to nonmusicians. High-arousal emotional speech (e.g., happiness, fear, and anger), often characterized by high pitch and/or large pitch variability, rapid speech rate, and/or increased intensity [18, 21], exhibits more pronounced shifts in these features. As a result, even slight increases in these features might be more salient to music subjects, potentially leading to higher intensity ratings. Furthermore, high-arousal emotional prosodies tend to capture more attention than low-arousal ones. Musicians typically demonstrate superior auditory attentional capacities relative to nonmusicians [169-171]. This heightened selective attention allows them to allocate more cognitive resources to processing emotional information in high-arousal speech prosody. This deeper processing amplifies the perceived intensity of the emotion. This dual effect of enhanced perception and attentional capacity may result in a heightened emotional experience, making high-arousal emotions seem more intense than those with low arousal.

Music majors in our study displayed similar levels of emotion recognition as nonmusic majors, but provided higher intensity ratings for emotional prosodies in Mandarin. This

dissociation pattern may arise from differences in task demands between emotion recognition and intensity ratings. Emotion recognition tasks, such as categorizing emotions, are generally regarded as explicit tasks; intensity rating tasks are seen as more implicit [172-174]. This distinction stems from the fact that categorizing emotions require direct access to a stable internal representation of a stimulus and a comparison of the presented stimulus with mental prototypes of each emotion. This process necessitates higher activation levels and a conscious representation of the stimulus. On the other hand, rating the intensity of an emotion may not require verbal categorization of a given emotion and relies on a weaker internal representation of the stimulus, with less precise knowledge about this stimulus being necessary for a more global judgement [173-175]. The dissociation between emotion categorization and emotion intensity judgment has been documented in the processing of musical [173, 175-177] and vocal [174, 178] emotion. Thus, our findings suggest that instrumental music training may primarily relate to the implicit, rather than explicit, processing of emotional prosody in tonal languages. This cross-domain transfer likely stems from a statistical learning mechanism employed in music and language acquisition. Statistical learning refers to a domain-general perceptual mechanism where learners unconsciously extract statistical patterns or regularities from sensory input over time, without explicit instruction [29, 179]. This type of learning is particularly relevant for implicit knowledge acquisition that relies on specific acoustic cues [180]. Therefore, even passive exposure to music allows both music and nonmusic subjects to acquire implicit knowledge about the statistical regularities between acoustic cues and musical emotions [181, 182]. On the basis of this implicit knowledge, all subjects are likely to show heightened attention to these cues in

emotional prosody and make intensity judgments. Nonetheless, music training enhances music subjects' sensitivity to the acoustic cues [183, 184], leading to a superior ability to attend to them compared to nonmusic subjects. This heightened sensitivity arises from the focused and intensive nature of music training, which makes music subjects particularly adept at detecting the underlying statistical regularities. Indeed, research demonstrates that musicians outperform nonmusicians in auditory statistical learning tasks involving pitch [161, 185-187] and rhythm [188-190]. Considering the implicit nature of statistical learning [191], its impact is more likely to be observed in implicit tasks rather than explicit ones.

This study has several limitations that warrant careful consideration. First, the quasi-experimental design precludes establishing a causal link between instrumental music training and emotional prosody processing. Future studies should employ randomized experimental designs to address this issue. Second, limited sample size likely contributed to the weakness of evidence for some findings. Consequently, subsequent future studies should use larger sample sizes to obtain more reliable results. Third, this investigation utilized only one tonal language, raising questions about the generalizability of the findings to other tonal languages. Future research is needed to explore this broader applicability. Finally, the mechanisms by which instrumental music training is related to the implicit processing of emotional prosody remain unclear and merit further exploration.

Conclusion

Despite the limitations outlined above, our findings do not support an association between instrumental music training and enhanced recognition of emotional prosody in tonal languages. However, they do suggest a link between music training and participants' tendency

to assign higher intensity ratings to emotional prosody. This dissociation between emotion recognition and emotion intensity evaluation underscores the complexity of the relationship between music training and emotion perception in tonal languages.

The present study is the first to investigate the relationship between music training and emotional prosody perception in Mandarin, a representative of tonal language. Future research should examine this question using different tonal languages to test the reliability of our results and deepen the understanding of the relationship. Such research has important implications for educators and clinicians, as it may provide a new method to treat impaired implicit processing of emotional prosody.

Supporting information

S1 Table. Overview of selected study characteristics for musically trained subjects.

S2 Table. Pairwise comparisons of emotional prosodies.

S3 Table. Pairwise comparisons of emotional prosodies. All statistics were obtained from the aligned-and-ranked data.

Acknowledgments

We are grateful to Dr. Pan Liu from the Department of Psychology at University of Alberta for providing us with the necessary experimental materials

Competing interests

There are no conflicts of interest.

Author contributions

Mengting Liu: Data curation, Funding acquisition, Investigation, Methodology, Resources, Writing—original draft, Writing—review & editing. **Xiangbin Teng:** Methodology, Writing—

original draft, Writing–review & editing. **Jun Jiang:** Conceptualization, Formal analysis, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review & editing.

References

1. Neves L, Martins M, Correia AI, Castro SL, Lima CF. Associations between vocal emotion recognition and socio-emotional adjustment in children. *Royal Society Open Science*. 2021;8(11). doi: doi:10.1098/rsos.211412.
2. Chronaki G, Wigelsworth M, Pell MD, Kotz SA. The development of cross-cultural recognition of vocal emotion during childhood and adolescence. *Scientific Reports*. 2018;8(1). doi: 10.1038/s41598-018-26889-1.
3. Alba-Ferrara L, Hausmann M, Mitchell RL, Weis S. The neural correlates of emotional prosody comprehension: Disentangling simple from complex emotion. *PLOS ONE*. 2011;6(12). doi: 10.1371/journal.pone.0028701.
4. Grandjean D. Brain networks of emotional prosody processing. *Emotion Review*. 2021;13(1):34-43. doi: 10.1177/1754073919898522.
5. Scherer KR, Banse R, Wallbott HG. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*. 2001;32(1):76-92. doi: 10.1177/0022022101032001009.
6. Young AW, Frühholz S, Schweinberger SR. Face and voice perception: Understanding commonalities and differences. *Trends in Cognitive Sciences*. 2020;24(5):398-410. doi: 10.1016/j.tics.2020.02.001.
7. Fridenson-Hayo S, Berggren S, Lassalle A, Tal S, Pigat D, Bölte S, et al. Basic and complex emotion recognition in children with autism: Cross-cultural findings. *Molecular Autism*. 2016;7(1). doi: 10.1186/s13229-016-0113-9.

8. Schelinski S, von Kriegstein K. The relation between vocal pitch and vocal emotion recognition abilities in people with autism spectrum disorder and typical development. *Journal of Autism and Developmental Disorders*. 2019;49(1):68-82. doi: 10.1007/s10803-018-3681-z.
9. Gong B, Li Q, Zhao Y, Wu C. Auditory emotion recognition deficits in schizophrenia: A systematic review and meta-analysis. *Asian Journal of Psychiatry*. 2021;65. doi: 10.1016/j.ajp.2021.102820.
10. Hoekert M, Kahn RS, Pijnenborg M, Aleman A. Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophrenia Research*. 2007;96(1):135-45. doi: 10.1016/j.schres.2007.07.023.
11. Icht M, Zukerman G, Ben-Itzhak E, Ben-David BM. Keep it simple: Identification of basic versus complex emotions in spoken language in individuals with autism spectrum disorder without intellectual disability: A meta-analysis study. *Autism Research*. 2021;14(9):1948-64. doi: 10.1002/aur.2551.
12. Pawełczyk A, Łojek E, Radek M, Pawełczyk T. Prosodic deficits and interpersonal difficulties in patients with schizophrenia. *Psychiatry Research*. 2021;306. doi: 10.1016/j.psychres.2021.114244.
13. Ong JH. Distributional learning of lexical tone and musical pitch by naïve and experienced adult learners [Doctoral dissertation]: Western Sydney University; 2016.
14. Lee C-Y, Tao L, Bond ZS. Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics*. 2008;36(4):537-63. doi: 10.1016/j.wocn.2008.01.002.

15. Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*. 1996;70(3):614-36. doi: 10.1037/0022-3514.70.3.614.
16. Johnstone T, Scherer KR. Vocal communication of emotion. In: Lewis M, Haviland-Jones JM, editors. *Handbook of emotions*. 2nd ed. New York, NY, US: Guilford Press; 2000. p. 220-35.
17. Goudbeek M, Scherer K. Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America*. 2010;128(3):1322-36. doi: 10.1121/1.3466853.
18. Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*. 2003;129(5):770-814. doi: 10.1037/0033-2909.129.5.770.
19. Juslin PN, Laukka P. Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*. 2001;1(4):381-412. doi: 10.1037/1528-3542.1.4.381.
20. Leitman DI, Laukka P, Juslin PN, Saccente E, Butler P, Javitt DC. Getting the cue: Sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophrenia Bulletin*. 2008;36(3):545-56. doi: 10.1093/schbul/sbn115.
21. Scherer KR. Vocal affect expression: A review and a model for future research. *Psychological Bulletin*. 1986;99(2):143-65. doi: 10.1037/0033-2909.99.2.143.
22. Curtis ME, Bharucha JJ. The minor third communicates sadness in speech, mirroring its use in music. *Emotion*. 2010;10(3):335-48. doi: 10.1037/a0017928.

23. Ilie G, Thompson WF. A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*. 2006;23(4):319-30. doi: 10.1525/mp.2006.23.4.319.
24. Coutinho E, Dibben N. Psychoacoustic cues to emotion in speech prosody and music. *Cognition and Emotion*. 2013;27(4):658-84. doi: 10.1080/02699931.2012.732559.
25. Thorndike EL, Woodworth RS. The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological Review*. 1901;8(3):247-61. doi: 10.1037/h0074898.
26. Patel AD. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*. 2011;2. doi: 10.3389/fpsyg.2011.00142.
27. Patel AD. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*. 2014;308:98-108. doi: 10.1016/j.heares.2013.08.011.
28. Besson M, Chobert J, Marie C. Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology*. 2011;2. doi: 10.3389/fpsyg.2011.00094.
29. Patel AD. *Music, Language, and the Brain*: Oxford University Press; 2008 22 Mar 2012.
30. Ong JH, Burnham D, Escudero P. Distributional learning of lexical tones: A comparison of attended vs. unattended listening. *PLOS ONE*. 2015;10(7). doi: 10.1371/journal.pone.0133446.

31. Asaridou SS, Hagoort P, McQueen JM. Effects of early bilingual experience with a tone and a non-tone language on speech-music integration. *PLOS ONE*. 2015;10(12). doi: 10.1371/journal.pone.0144225.
32. Chen A, Peter V, Wijnen F, Schnack H, Burnham D. Are lexical tones musical? Native language's influence on neural response to pitch in different domains. *Brain and Language*. 2018;180-182:31-41. doi: 10.1016/j.bandl.2018.04.006.
33. Creel SC. The familiar-melody advantage in auditory perceptual development: Parallels between spoken language acquisition and general auditory perception. *Attention, Perception, & Psychophysics*. 2019;81(4):948-57. doi: 10.3758/s13414-018-01663-7.
34. Creel SC, Obiri-Yeboah M, Rose S. Language-to-music transfer effects depend on the tone language: Akan vs. East Asian tone languages. *Memory & Cognition*. 2023;51(7):1624-39. doi: 10.3758/s13421-023-01416-4.
35. Ip MHK, Cutler A. Universals of listening: Equivalent prosodic entrainment in tone and non-tone languages. *Cognition*. 2020;202. doi: 10.1016/j.cognition.2020.104311.
36. Peng G, Zheng H-Y, Gong T, Yang R-X, Kong J-P, Wang WSY. The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*. 2010;38(4):616-24. doi: 10.1016/j.wocn.2010.09.003.
37. Ong JH, Wong PCM, Liu F. Musicians show enhanced perception, but not production, of native lexical tones. *The Journal of the Acoustical Society of America*. 2020;148(6):3443-54. doi: 10.1121/10.0002776.

38. Correia AI, Castro SL, MacGregor C, Müllensiefen D, Schellenberg EG, Lima CF. Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*. 2022;22(5):894-906. doi: 10.1037/emo0000770.
39. Farmer E, Jicol C, Petrini K. Musicianship enhances perception but not feeling of emotion from others' social interaction through speech prosody. *Music Perception*. 2020;37(4):323-38. doi: 10.1525/mp.2020.37.4.323.
40. Lima CF, Castro SL. Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*. 2011;11(5):1021-31. doi: 10.1037/a0024521.
41. Twaite JT. Examining relationships between basic emotion perception and musical training in the prosodic, facial, and lexical channels of communication and in music [Doctoral dissertation]: City University of New York; 2016.
42. Martins M, Pinheiro AP, Lima CF. Does music training improve emotion recognition abilities? A critical review. *Emotion Review*. 2021;13(3):199-210. doi: 10.1177/17540739211022035.
43. Schellenberg EG, Lima CF. Music training and nonmusical abilities. *Annual Review of Psychology*. 2024;75:87-128. doi: 10.1146/annurev-psych-032323-051354.
44. Nussbaum C, Schweinberger SR. Links between musicality and vocal emotion perception. *Emotion Review*. 2021;13(3):211-24. doi: 10.1177/17540739211022803.
45. Thompson WF, Schellenberg EG, Husain G. Perceiving prosody in speech. *Annals of the New York Academy of Sciences*. 2003;999(1):530-2. doi: 10.1196/annals.1284.067.

46. Thompson WF, Schellenberg EG, Husain G. Decoding speech prosody: Do music lessons help? *Emotion*. 2004;4(1):46-64. doi: 10.1037/1528-3542.4.1.46.
47. Fuller CD, Galvin JJ, Maat B, Free RH, Başkent D. The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience*. 2014;8. doi: 10.3389/fnins.2014.00179.
48. Nussbaum C, Schirmer A, Schweinberger SR. Musicality – Tuned to the melody of vocal emotions. *British Journal of Psychology*. 2023. doi: 10.1111/bjop.12684.
49. Schellenberg EG, Mankarious M. Music training and emotion comprehension in childhood. *Emotion*. 2012;12(5):887-91. doi: 10.1037/a0027971.
50. Francisco RMC. Pianos and microphones: Does the type of musical training affect emotion recognition? [Master's thesis]: Universidade de Lisboa; 2021.
51. Başkent D, Fuller CD, Galvin JJ, III, Schepel L, Gaudrain E, Free RH. Musician effect on perception of spectro-temporally degraded speech, vocal emotion, and music in young adolescents. *The Journal of the Acoustical Society of America*. 2018;143(5):EL311-EL6. doi: 10.1121/1.5034489.
52. Mualem O, Lavidor M. Music education intervention improves vocal emotion recognition. *International Journal of Music Education*. 2015;33(4):413-25. doi: 10.1177/0255761415584292.
53. Park M, Gutyrchik E, Welker L, Carl P, Pöppel E, Zaytseva Y, et al. Sadness is unique: Neural processing of emotions in speech prosody in musicians and non-musicians. *Frontiers in Human Neuroscience*. 2015;8. doi: 10.3389/fnhum.2014.01049.

54. Zhang JD, Susino M, McPherson GE, Schubert E. The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music*. 2020;48(3):389-409. doi: 10.1177/0305735618804038.
55. Gómez-de-Mariscal E, Guerrero V, Sneider A, Jayatilaka H, Phillip JM, Wirtz D, et al. Use of the p-values as a size-dependent function to address practical differences when analyzing large datasets. *Scientific Reports*. 2021;11(1). doi: 10.1038/s41598-021-00199-5.
56. Whitley E, Ball J. Statistics review 4: Sample size calculations. *Critical Care*. 2002;6(4):335-41. doi: 10.1186/cc1521.
57. Brown S. The "musilanguage" model of music evolution. In: Wallin NL, Merker B, Brown S, editors. *The origins of music*: MIT Press; 1999. p. 271-300.
58. Burnham D, Jones C. Categorical perception of lexical tone by tonal and non-tonal language speakers. 9th International Conference on Speech Science and Technology; Melbourne, VIC, Australia 2002, December 2–5. p. 515-20.
59. Hallé PA, Chang Y-C, Best CT. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*. 2004;32(3):395-421. doi: 10.1016/S0095-4470(03)00016-0.
60. Kitayama S, Ishii K. Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion*. 2002;16(1):29-59. doi: 10.1080/0269993943000121.

61. Ishii K, Reyes JA, Kitayama S. Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science*. 2003;14(1):39-46. doi: 10.1111/1467-9280.01416. PubMed PMID: 12564752.
62. Yang Y, Wang L, Wang Q. Take your word or tone for it? European American and Chinese children's attention to emotional cues in speech. *Child Development*. 2021;92(3):844-52. doi: 10.1111/cdev.13576.
63. Good A, Gordon KA, Papsin BC, Nespoli G, Hopyan T, Peretz I, et al. Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants. *Ear and Hearing*. 2017;38(4). doi: 10.1097/aud.0000000000000402.
64. Bodner E, Aharoni R, Iancu I. The effect of training with music on happiness recognition in social anxiety disorder. *Journal of Psychopathology and Behavioral Assessment*. 2012;34(4):458-66. doi: 10.1007/s10862-012-9304-7.
65. Chari DA, Barrett KC, Patel AD, Colgrove TR, Jiradejvong P, Jacobs LY, et al. Impact of auditory-motor musical training on melodic pattern recognition in cochlear implant users. *Otology & Neurotology*. 2020;41(4):e422-e31. doi: 10.1097/MAO.0000000000002525.
66. Fuller CD, Galvin JJ, Maat B, Başkent D, Free RH. Comparison of two music training approaches on music and speech perception in cochlear implant users. *Trends in Hearing*. 2018;22. doi: 10.1177/2331216518765379. PubMed PMID: 29621947.
67. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*. 2007;39(2):175-91. doi: 10.3758/BF03193146.

68. Zhang JD, Schubert E. A single item measure for identifying musician and nonmusician categories based on measures of musical sophistication. *Music Perception*. 2019;36(5):457-67. doi: 10.1525/mp.2019.36.5.457.
69. Müllensiefen D, Gingras B, Musil J, Stewart L. The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE*. 2014;9(2). doi: 10.1371/journal.pone.0089642.
70. Chierchia G, Fuhrmann D, Knoll LJ, Pi-Sunyer BP, Sakhardande AL, Blakemore S-J. The matrix reasoning item bank (MaRs-IB): Novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science*. 2019;6(10). doi: 10.1098/rsos.190232.
71. Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*. 1971;9(1):97-113. doi: 10.1016/0028-3932(71)90067-4.
72. Liu P, Pell MD. Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods*. 2012;44(4):1042-51. doi: 10.3758/s13428-012-0203-3.
73. Pell MD, Jaywant A, Monetta L, Kotz SA. Emotional speech processing: Disentangling the effects of prosody and semantic cues. *Cognition and Emotion*. 2011;25(5):834-53. doi: 10.1080/02699931.2010.516915.
74. Larrouy-Maestri P, Poeppel D, Pell MD. The sound of emotional prosody: Nearly 3 decades of research and future directions. *Perspectives on Psychological Science*. 2024. doi: 10.1177/17456916231217722. PubMed PMID: 38232303.

75. Castro SL, Lima CF. Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods*. 2010;42(1):74-81. doi: 10.3758/BRM.42.1.74.
76. Liu P, Rigoulot S, Jiang X, Zhang S, Pell MD. Unattended emotional prosody affects visual processing of facial expressions in Mandarin-speaking Chinese: A comparison with English-speaking Canadians. *Journal of Cross-Cultural Psychology*. 2021;52(3):275-94. doi: 10.1177/0022022121990897.
77. Paulmann S, Uskul AK. Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition and Emotion*. 2014;28(2):230-44. doi: 10.1080/02699931.2013.812033.
78. Raven J. The Raven Progressive Matrices and measuring aptitude constructs. *The International Journal of Educational and Psychological Assessment*. 2009;2:2-38. PubMed PMID: 2010-11467-001.
79. Zorowitz S, Chierchia G, Blakemore S-J, Daw ND. An item response theory analysis of the matrix reasoning item bank (MaRs-IB). *Behavior Research Methods*. 2023. doi: 10.3758/s13428-023-02067-8.
80. Li J, Lin H-R, Wolf A, Lothwesen K. Measuring musical sophistication in the Chinese general population: Validation and replication of the Simplified Chinese Gold-MSI. *Musicae Scientiae*. 2023. doi: 10.1177/10298649231183264.
81. Lin H-R, Kopiez R, Müllensiefen D, Wolf A. The Chinese version of the Gold-MSI: Adaptation and validation of an inventory for the measurement of musical sophistication

- in a Taiwanese sample. *Musicae Scientiae*. 2019;25(2):226-51. doi: 10.1177/1029864919871987.
82. Santos AO, Godoy J, Silverio K, Brasolotto A. Vocal changes of men and women from different age decades: An analysis from 30 years of age. *Journal of Voice*. 2023;37(6):840-50. doi: 10.1016/j.jvoice.2021.06.003.
 83. Hillenbrand JM, Clark MJ. The role of f_0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*. 2009;71(5):1150-66. doi: 10.3758/APP.71.5.1150.
 84. Heitz RP. The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*. 2014;8. doi: 10.3389/fnins.2014.00150.
 85. Wickelgren WA. Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*. 1977;41(1):67-85. doi: 10.1016/0001-6918(77)90012-9.
 86. Alhowaide A, Alsmadi I, Tang J. Ensemble detection model for IoT IDS. *Internet of Things*. 2021;16. doi: 10.1016/j.iot.2021.100435.
 87. Lerche V, Voss A. Speed–accuracy manipulations and diffusion modeling: Lack of discriminant validity of the manipulation or of the parameter estimates? *Behavior Research Methods*. 2018;50(6):2568-85. doi: 10.3758/s13428-018-1034-7.
 88. Lerche V, Voss A. When accuracy rates and mean response times lead to false conclusions: A simulation study based on the diffusion model. *TQMP*. 2020;16(2):107-19. doi: 10.20982/tqmp.16.2.p107.
 89. Mueller ST, Alam L, Funke GJ, Linja A, Ibne Mamun T, Smith SL. Examining methods for combining speed and accuracy in a go/no-go vigilance task. *Proceedings of the Human*

- Factors and Ergonomics Society Annual Meeting. 2020;64(1):1202-6. doi: 10.1177/1071181320641286.
90. Bakun Emesh T, Garbi D, Kaplan A, Zelicha H, Yaskolka Meir A, Tsaban G, et al. Retest reliability of integrated speed–accuracy measures. *Assessment*. 2022;29(4):717-30. doi: 10.1177/1073191120985609. PubMed PMID: 33522278.
 91. Liesefeld HR, Janczyk M. Combining speed and accuracy to control for speed-accuracy trade-offs(?). *Behavior Research Methods*. 2019;51(1):40-60. doi: 10.3758/s13428-018-1076-x.
 92. Liesefeld HR, Janczyk M. Same same but different: Subtle but consequential differences between two measures to linearly integrate speed and accuracy (LISAS vs. BIS). *Behavior Research Methods*. 2023;55(3):1175-92. doi: 10.3758/s13428-022-01843-2.
 93. Liesefeld HR, Fu X, Zimmer HD. Fast and careless or careful and slow? Apparent holistic processing in mental rotation is explained by speed-accuracy trade-offs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2015;41(4):1140-51. doi: 10.1037/xlm0000081.
 94. Bouvet L, Simard-Meilleur A-A, Paignon A, Motttron L, Donnadieu S. Auditory local bias and reduced global interference in autism. *Cognition*. 2014;131(3):367-72. doi: 10.1016/j.cognition.2014.02.006.
 95. Hassan EK, Jones AM, Buckingham G. A novel protocol to induce mental fatigue. *Behavior Research Methods*. 2023. doi: 10.3758/s13428-023-02191-5.

96. Paas Oliveros LK, Cieslik EC, Pieczykolan A, Pläschke RN, Eickhoff SB, Langner R. Brain functional characterization of response-code conflict in dual-tasking and its modulation by age. *Cerebral Cortex*. 2023;33(18):10155-80. doi: 10.1093/cercor/bhad273.
97. English MCW, Maybery MT, Visser TAW. Magnitude of sex differences in visual search varies with target eccentricity. *Psychonomic Bulletin & Review*. 2021;28(1):178-88. doi: 10.3758/s13423-020-01796-7.
98. Weissbach A, Moyé J, Takacs A, Verrel J, Chwolka F, Friedrich J, et al. Perception–action integration is altered in functional movement disorders. *Movement Disorders*. 2023;38(8):1399-409. doi: 10.1002/mds.29458.
99. Robson SG, Tangen JM, Searston RA. The effect of expertise, target usefulness and image structure on visual search. *Cognitive Research: Principles and Implications*. 2021;6. doi: 10.1186/s41235-021-00282-5.
100. Yu Z, Dong Y. The emergence of a complex language skill: Evidence from the self-organization of interpreting competence in interpreting students. *Bilingualism: Language and Cognition*. 2022;25(2):269-82. Epub 2021/09/28. doi: 10.1017/S1366728921000870.
101. McGarrigle R, Knight S, Hornsby BWY, Mattys S. Predictors of listening-related fatigue across the adult life span. *Psychological Science*. 2021;32(12):1937-51. doi: 10.1177/09567976211016410. PubMed PMID: 34751602.
102. McGarrigle R, Mattys S. Sensory-processing sensitivity predicts fatigue from listening, but not perceived effort, in young and older adults. *Journal of Speech, Language, and Hearing Research*. 2023;66(2):444-60. doi: doi:10.1044/2022_JSLHR-22-00374.

103. Calin-Jageman RJ, Cumming G. Estimation for better inference in neuroscience. *eNeuro*. 2019;6(4). doi: 10.1523/eneuro.0205-19.2019.
104. Calin-Jageman RJ, Cumming G. The new statistics for better science: Ask how much, how uncertain, and what else is known. *The American Statistician*. 2019;73(sup1):271-80. doi: 10.1080/00031305.2018.1518266.
105. Cohen J. Things I have learned (so far). *American Psychologist*. 1990;45(12):1304-12. doi: 10.1037/0003-066X.45.12.1304.
106. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*. 2007;82(4):591-605. doi: 10.1111/j.1469-185X.2007.00027.x.
107. Wasserstein RL, Lazar NA. The ASA statement on p-values: Context, process, and purpose. *The American Statistician*. 2016;70(2):129-33. doi: 10.1080/00031305.2016.1154108.
108. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician*. 2019;73(sup1):1-19. doi: 10.1080/00031305.2019.1583913.
109. American Psychological Association. Publication manual of the American Psychological Association. 6th ed2010.
110. American Psychological Association. Publication manual of the American Psychological Association. 7th ed2020.
111. Points of significance. *Nature Human Behaviour*. 2023;7(3):293-4. doi: 10.1038/s41562-023-01586-w.

112. Harrington D, D'Agostino RB, Gatsonis C, Hogan JW, Hunter DJ, Normand S-LT, et al. New guidelines for statistical reporting in the Journal. *The New England Journal of Medicine*. 2019;381(3):285-6. doi: 10.1056/NEJMe1906559. PubMed PMID: 31314974.
113. Trafimow D, Marks M. Editorial. *Basic and Applied Social Psychology*. 2015;37(1):1-2. doi: 10.1080/01973533.2015.1012991.
114. Cumming G. The new statistics: A how-to guide. *Australian Psychologist*. 2013;48(3):161-70. doi: 10.1111/ap.12018.
115. Cumming G. The new statistics: Why and how. *Psychological Science*. 2014;25(1):7-29. doi: 10.1177/0956797613504966. PubMed PMID: 24220629.
116. Griffiths P, Needleman J. Statistical significance testing and p-values: Defending the indefensible? A discussion paper and position statement. *International Journal of Nursing Studies*. 2019;99. doi: 10.1016/j.ijnurstu.2019.07.001.
117. Karadaghy OA, Hong H, Scott-Wittenborn N, Sinha P, Suko J, Tait S, et al. Reporting of effect size and confidence intervals in JAMA Otolaryngology–Head & Neck Surgery. *JAMA Otolaryngology–Head & Neck Surgery*. 2017;143(11):1075-80. doi: 10.1001/jamaoto.2017.1504.
118. Cheung SF, Pesigan IJA, Vong WN. DIY bootstrapping: Getting the nonparametric bootstrap confidence interval in SPSS for any statistics or function of statistics (when this bootstrapping is appropriate). *Behavior Research Methods*. 2023;55(2):474-90. doi: 10.3758/s13428-022-01808-5.

119. Perdices M. Null hypothesis significance testing, p-values, effects sizes and confidence intervals. *Brain Impairment*. 2018;19(1):70-80. Epub 2017/12/07. doi: 10.1017/BrImp.2017.28.
120. Sohn D. Sex differences in achievement self-attributions: An effect-size analysis. *Sex Roles*. 1982;8(4):345-57. doi: 10.1007/BF00287274.
121. Steiger JH. Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*. 2004;9(2):164-82. doi: 10.1037/1082-989X.9.2.164.
122. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
123. Cohen J. A power primer. *Psychological Bulletin*. 1992;112(1):155-9. doi: 10.1037/0033-2909.112.1.155.
124. Kirk RE. Practical significance: A concept whose time has come. *Educational and Psychological Measurement*. 1996;56(5):746-59. doi: 10.1177/0013164496056005002.
125. Delacre M, Lakens D, Ley C, Liu L, Leys C. Why Hedges' g 's based on the non-pooled standard deviation should be reported with Welch's t -test2021.
126. Bakdash JZ, Marusich LR. Repeated measures correlation. *Frontiers in Psychology*. 2017;8. doi: 10.3389/fpsyg.2017.00456.
127. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 1—correlation within subjects. *BMJ*. 1995;310(6977). doi: 10.1136/bmj.310.6977.446.

128. Marusich L, Bakdash J. rmcrrShiny: A web and standalone application for repeated measures correlation. *F1000Research*. 2021;10. doi: 10.12688/f1000research.55027.2.
129. Chihara LM, Hesterberg TC. *Mathematical statistics with resampling and R*. 3rd ed: John Wiley & Sons; 2022.
130. Hair JF, Hult GTM, Ringle CM, Sarstedt M. *A primer on partial least squares structural equation modeling (PLS-SEM)*. 3rd ed: SAGE Publications; 2022.
131. Streukens S, Leroi-Werelds S. Bootstrapping and PLS-SEM: A step-by-step guide to get more out of your bootstrap results. *European Management Journal*. 2016;34(6):618-32. doi: 10.1016/j.emj.2016.06.003.
132. Hesterberg TC. What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*. 2015;69(4):371-86. doi: 10.1080/00031305.2015.1089789.
133. Brydges CR, Bielak AAM. A Bayesian analysis of evidence in support of the null hypothesis in gerontological psychology (or lack thereof). *The Journals of Gerontology: Series B*. 2020;75(1):58-66. doi: 10.1093/geronb/gbz033.
134. Heck DW, Boehm U, Böing-Messing F, Bürkner P-C, Derks K, Dienes Z, et al. A review of applications of the Bayes factor in psychological research. *Psychological Methods*. 2023;28(3):558-79. doi: 10.1037/met0000454.
135. Schmalz X, Biurrun Manresa J, Zhang L. What is a Bayes factor? *Psychological Methods*. 2023;28(3):705-18. doi: 10.1037/met0000421.

136. Wagenmakers E-J, Marsman M, Jamil T, Ly A, Verhagen J, Love J, et al. Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*. 2018;25(1):35-57. doi: 10.3758/s13423-017-1343-3.
137. Jeffreys H. *Theory of probability*. 3rd ed. Oxford, UK: Oxford University Press; 1961.
138. Lee MD, Wagenmakers E-J. *Bayesian cognitive modeling: A practical course*. New York, NY: Cambridge University Press; 2014. xiii, 264-xiii, p.
139. Fu Q, Hoijtink H, Moerbeek M. Sample-size determination for the Bayesian t test and Welch's test using the approximate adjusted fractional Bayes factor. *Behavior Research Methods*. 2021;53(1):139-52. doi: 10.3758/s13428-020-01408-1.
140. Jamil T, Ly A, Morey RD, Love J, Marsman M, Wagenmakers E-J. Default "Guel and Dickey" Bayes factors for contingency tables. *Behavior Research Methods*. 2017;49(2):638-52. doi: 10.3758/s13428-016-0739-8.
141. Wayland R, Herrera E, Kaan E. Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*. 2010;38(4):654-62. doi: 10.1016/j.wocn.2010.10.001.
142. Wong PCM, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*. 2007;10(4):420-2. doi: 10.1038/nn1872.
143. Alexander J, Wong P, Bradlow A. Lexical tone perception in musicians and non-musicians. 6th Interspeech 2005 and 9th European Conference on Speech Communication and Technology; Lisbon, Portugal: International Speech Communication Association; 2005, September 4–8. p. 397-400.

144. Cooper A, Wang Y. The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America*. 2012;131(6):4756-69. doi: 10.1121/1.4714355.
145. Lee C-Y, Lekich A, Zhang Y. Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*. 2014;135(3):1607-15. doi: 10.1121/1.4864473.
146. Chen S, Zhu Y, Wayland R, Yang Y. How musical experience affects tone perception efficiency by musicians of tonal and non-tonal speakers? *PLOS ONE*. 2020;15(5). doi: 10.1371/journal.pone.0232514.
147. Mok PKP, Zuo D. The separation between music and speech: Evidence from the perception of Cantonese tones. *The Journal of the Acoustical Society of America*. 2012;132(4):2711-20. doi: 10.1121/1.4747010.
148. Maggu AR, Wong PCM, Liu H, Wong FCK. Experience-dependent influence of music and language on lexical pitch learning is not additive. 19th Annual Conference of the International Speech Communication Association (Interspeech 2018); Hyderabad, India. Hyderabad, India 2018, September 2–6. p. 3791-4.
149. Deutsch D, Henthorn T, Marvin E, Xu H. Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America*. 2006;119(2):719-22. doi: 10.1121/1.2151799.

150. Deutsch D, Li X, Shen J. Absolute pitch among students at the Shanghai Conservatory of Music: A large-scale direct-test study. *The Journal of the Acoustical Society of America*. 2013;134(5):3853-9. doi: 10.1121/1.4824450.
151. Miyazaki Ki, Rakowski A, Makomaska S, Jiang C, Tsuzaki M, Oxenham AJ, et al. Absolute pitch and relative pitch in music students in the East and the West: Implications for aural-skills education. *Music Perception*. 2018;36(2):135-55. doi: 10.1525/mp.2018.36.2.135.
152. Lee C-Y, Lee Y-F. Perception of musical pitch and lexical tones by Mandarin-speaking musicians. *The Journal of the Acoustical Society of America*. 2010;127(1):481-90. doi: 10.1121/1.3266683.
153. Pell MD, Paulmann S, Dara C, Alasseri A, Kotz SA. Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*. 2009;37(4):417-35. doi: 10.1016/j.wocn.2009.07.005.
154. Breitenstein C, Lancker DV, Daum I. The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion*. 2001;15(1):57-79. doi: 10.1080/02699930126095.
155. Miyazaki Ki, Makomaska S, Rakowski A. Prevalence of absolute pitch: A comparison between Japanese and Polish music students. *The Journal of the Acoustical Society of America*. 2012;132(5):3484-93. doi: 10.1121/1.4756956.
156. Leite RBC, Mota-Rolim SA, Queiroz CMT. Music proficiency and quantification of absolute pitch: A large-scale study among Brazilian musicians. *Frontiers in Neuroscience*. 2016;10. doi: 10.3389/fnins.2016.00447.

157. Ngo MK, Vu K-PL, Strybel TZ. Effects of music and tonal language experience on relative pitch performance. *The American Journal of Psychology*. 2016;129(2):125-34. doi: 10.5406/amerjpsyc.129.2.0125.
158. Schellenberg EG, Moreno S. Music lessons, pitch processing, and g. *Psychology of Music*. 2010;38(2):209-21. doi: 10.1177/0305735609339473.
159. Faber S, Fiveash A. Emotion without words: A comparison study of music and speech prosody. *Canadian Journal of Music Therapy*. 2014;20(2):86-101.
160. Dibben N, Coutinho E, Vilar JA, Estévez-Pérez G. Do individual differences influence moment-by-moment reports of emotion perceived in music and speech prosody? *Frontiers in Behavioral Neuroscience*. 2018;12. doi: 10.3389/fnbeh.2018.00184.
161. Vasuki PRM, Sharma M, Demuth K, Arciuli J. Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research*. 2016;342:112-23. doi: 10.1016/j.heares.2016.10.008.
162. Micheyl C, Delhommeau K, Perrot X, Oxenham AJ. Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*. 2006;219(1):36-47. doi: 10.1016/j.heares.2006.05.004.
163. Liang C, Earl B, Thompson I, Whitaker K, Cahn S, Xiang J, et al. Musicians are better than non-musicians in frequency change detection: behavioral and electrophysiological evidence. *Frontiers in Neuroscience*. 2016;10. doi: 10.3389/fnins.2016.00464.
164. Kishon-Rabin L, Amir O, Vexler Y, Zaltz Y. Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology*. 2001;12(2):125-44. doi: doi:10.1515/JBCPP.2001.12.2.125.

165. Sares AG, Foster NEV, Allen K, Hyde KL. Pitch and time processing in speech and tones: The effects of musical training and attention. *Journal of Speech, Language, and Hearing Research*. 2018;61(3):496-509. doi: 10.1044/2017_JSLHR-S-17-0207.
166. Slade T, Gascon A, Comeau G, Russell D. Just noticeable differences in sound intensity of piano tones in non-musicians and experienced pianists. *Psychology of Music*. 2023;51(3):924-37. doi: 10.1177/03057356221126203.
167. Madsen CK. Modulated beat discrimination among musicians and nonmusicians. *Journal of Research in Music Education*. 1979;27(2):57-67. doi: 10.2307/3344892.
168. Nguyen T, Sidhu RK, Everling JC, Wickett MC, Gibbings A, Grahn JA. Beat perception and production in musicians and dancers. *Music Perception*. 2022;39(3):229-48. doi: 10.1525/mp.2022.39.3.229.
169. Strait DL, Kraus N, Parbery-Clark A, Ashley R. Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*. 2010;261(1):22-9. doi: 10.1016/j.heares.2009.12.021.
170. Weijkamp J, Sadakata M. Attention to affective audio-visual information: Comparison between musicians and non-musicians. *Psychology of Music*. 2017;45(2):204-15. doi: 10.1177/0305735616654216.
171. Bialystok E, DePape A-M. Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*. 2009;35(2):565-74. doi: 10.1037/a0012735.
172. Cunningham WA, Raye CL, Johnson MK. Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes.

- Journal of Cognitive Neuroscience. 2004;16(10):1717-29. doi: 10.1162/0898929042947919.
173. Lévêque Y, Teyssier P, Bouchet P, Bigand E, Caclin A, Tillmann B. Musical emotions in congenital amusia: Impaired recognition, but preserved emotional intensity. *Neuropsychology*. 2018;32(7):880-94. doi: 10.1037/neu0000461.
174. Pralus A, Fornoni L, Bouet R, Gomot M, Bhatara A, Tillmann B, et al. Emotional prosody in congenital amusia: Impaired and spared processes. *Neuropsychologia*. 2019;134. doi: 10.1016/j.neuropsychologia.2019.107234.
175. Pralus A, Belfi A, Hirel C, Lévêque Y, Fornoni L, Bigand E, et al. Recognition of musical emotions and their perceived intensity after unilateral brain damage. *Cortex*. 2020;130:78-93. doi: 10.1016/j.cortex.2020.05.015.
176. Hirel C, Lévêque Y, Deiana G, Richard N, Cho TH, Mechtouff L, et al. Amusie acquise et anhédonie musicale. *Revue Neurologique*. 2014;170(8):536-40. doi: 10.1016/j.neurol.2014.03.015.
177. Lévêque Y, Roulet-Perez E, Deonna T, Moulin A, Fornoni L, Mayor-Dubois C, et al. Music processing deficits in Landau-Kleffner syndrome: Four case studies in adulthood. *Cortex*. 2020;129:99-111. doi: 10.1016/j.cortex.2020.03.025.
178. Bourgeois-Vionnet J, Moulin A, Hermier M, Pralus A, Nighoghossian N. A case of verbal and emotional prosody processing dissociation after a right temporal venous infarct. *Neurological Sciences*. 2020;41(6):1615-8. doi: 10.1007/s10072-019-04175-w.

179. Frost R, Armstrong BC, Siegelman N, Christiansen MH. Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*. 2015;19(3):117-25. doi: 10.1016/j.tics.2014.12.010.
180. Ong JH, Burnham D, Stevens CJ, Escudero P. Naïve learners show cross-domain transfer after distributional learning: The case of lexical and musical pitch. *Frontiers in Psychology*. 2016;7. doi: 10.3389/fpsyg.2016.01189.
181. Pearce MT. Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*. 2018;1423(1):378-95. doi: <https://doi.org/10.1111/nyas.13654>.
182. Bigand E, Poulin-Charronnat B. Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*. 2006;100(1):100-30. doi: 10.1016/j.cognition.2005.11.007.
183. François C, Schön D. Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: The role of musical practice. *Hearing Research*. 2014;308:122-8. doi: 10.1016/j.heares.2013.08.018.
184. Schön D, François C. Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology*. 2011;2. doi: 10.3389/fpsyg.2011.00167.
185. Vasuki PRM, Sharma M, Ibrahim RK, Arciuli J. Musicians’ Online Performance during Auditory and Visual Statistical Learning Tasks. *Frontiers in Human Neuroscience*. 2017;11. doi: 10.3389/fnhum.2017.00114.

186. Paraskevopoulos E, Kuchenbuch A, Herholz SC, Pantev C. Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia*. 2012;50(2):341-9. doi: 10.1016/j.neuropsychologia.2011.12.007.
187. François C, Schön D. Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*. 2011;21(10):2357-65.
188. Daikoku T, Yumoto M. Musical expertise facilitates statistical learning of rhythm and the perceptive uncertainty: A cross-cultural study. *Neuropsychologia*. 2020;146. doi: 10.1016/j.neuropsychologia.2020.107553.
189. van Zuijen TL, Sussman E, Winkler I, Näätänen R, Tervaniemi M. Auditory organization of sound sequences by a temporal or numerical regularity—a mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research*. 2005;23(2):270-6. doi: 10.1016/j.cogbrainres.2004.10.007.
190. Vasuki PRM, Sharma M, Ibrahim R, Arciuli J. Statistical learning and auditory processing in children with music training: An ERP study. *Clinical Neurophysiology*. 2017;128(7):1270-81. doi: 10.1016/j.clinph.2017.04.010.
191. Perruchet P, Pacton S. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*. 2006;10(5):233-8. doi: 10.1016/j.tics.2006.03.006.