Visual Congruency Modulates Music Reward through Sensorimotor Integration

Lei Zhang[1,2], Yi Du[1,3,6*], Robert J. Zatorre[2,4,5*]

[1] Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China

[2] Cognitive Neuroscience Unit, Montreal Neurological Institute, McGill University, Montreal, QC H3A 2B4, Canada.

[3] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China 100049

[4] Centre for Research in Brain, Language, and Music (CRBLM), McGill University, Faculty of Medicine 3640 de la Montagne, Montreal QC H3G 2A8, Canada

[5] International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, QC H2V 2J2, Canada.

[6] Chinese Institute for Brain Research, Beijing, China 102206

* Corresponding author:

Robert J. Zatorre, Ph.D.

3801 Rue University, Montréal, QC H3A 2B4

Email: robert.zatorre@mcgill.ca

Yi Du, Ph.D.

16 Lincui Road, Chaoyang district, Beijing, China 100101

Email: duyi@psych.ac.cn

Abstract

There is emerging evidence that a performer's body movements may enhance the music-induced pleasure of audiences. However, the neural mechanism underlying such modulation remains largely unexplored. This study utilized psychophysiological and electroencephalographic data collected from listeners as they watched and listened to manipulated vocal (Mandarin lyrics) and violin performances of Japanese and Chinese pop music. All participants were unfamiliar with the violin or Mandarin. The auditory and visual elements of the stimuli were either congruent (original recording) or incongruent (drawn from unrelated music videos). We found that congruent visual movements, as opposed to incongruent ones, increased both subjective pleasure ratings and skin conductance responses but only during vocal performances. Then, we examined the coherence between the music signal and sensorimotor Mu-band oscillatory neural activity and found that congruent visual movements enhanced Mu entrainment exclusively to vocal music signal. Further, mediation analysis demonstrated that neural entrainment to vocal music significantly mediated the visual modulation of music-induced pleasure. In conclusion, our study provides novel evidence on how congruent visual movements can heighten music-induced pleasure through enhanced sensorimotor integration.

Introduction

Music consists of acoustic sequences that can give us pleasure (Zatorre, 2023). However, music often transcends purely auditory art; enjoyment frequently arises from both the auditory experience and the visual observation of performers' gestures, whether live or via video. In addition to the audio part of music, performers' body movements and facial expressions are also a part of their performance. While the audio aspects of music are well-recognized, the role of information conveyed visually in enhancing music-induced pleasure remains poorly understood.

Prior behavioral research has explored the impact of visual movement cues on music enjoyment. A meta-analysis of 15 studies quantified the average effect size of the enhancement in music performance appreciation under audio-visual conditions compared to audio-only conditions. It found a relatively consistent medium effect size increase in pleasure components was present (Platz and Kopiez, 2012). Moreover, several studies have suggested that visual cues may have a more significant influence on judgments about music performance than auditory cues (Griffiths and Reay, 2018; Tsay, 2013). These findings imply that the visual elements of music play a role in the experience of pleasure from music, at least at the behavioral level. However, further objective evidence is needed to substantiate these findings and to understand the potential physiological and neural mechanisms.

Skin conductance response (SCR), driven by sympathetic nervous system, measures phasic changes in the electrical conductivity of the skin (Braithwaite et al., 2013). SCR is often used as an objective indication of physiological arousal, and can be modulated by music-induced pleasure. When subjects hear music that they rate as pleasant, SCRs increase with increasing pleasurable states (Mas-Herrero et al., 2014; Salimpoor et al., 2009). However, no study has yet examined whether observing performers' gestures could enhance the pleasure derived from music using objective physiological measures.

The enjoyment of music is also associated with movements in synchrony with the rhythm of the music. The shared affective motion experience (SAME) model proposes that music is perceived not only as an auditory signal but also through the intentional, hierarchically organized sequences of expressive motor acts behind the signal (Overy and Molnar-Szakacs, 2009). While listening to music, listeners may mimic the articulation or other movements of the performer, either overtly or covertly, facilitating the corepresentation and sharing of a musical experience between performer and listener (Dalla Bella et al., 2001; Gagnon and Peretz, 2003; Gomez and Danuser, 2007; Michalak et al., 2009; Overy and Molnar-Szakacs, 2009; Sievers et al., 2013). Neuroimaging studies have highlighted the engagement of the motor system during music listening, especially when the listener is proficient with the instrument being played (Gordon et al., 2018; Herholz et al., 2016; Lahav et al., 2007; Srinivasan et al., 2020), indicating that the motor system may play a role in aspects of perception.

Mu wave activity, a brain rhythm typically defined in the 8–13 Hz frequency band, provides a valid approach for studying the human motor system and sensorimotor integration with electroencephalogram (EEG) (Fox et al., 2016; Hobson and Bishop, 2017; Jenson et al., 2020). Reduced spectral power in the mu wave range, which originates in premotor areas and inferior parietal lobule, is observed during both action execution and action observation (Bonini et al., 2022; Fox et al., 2016, 2016; Hobson and Bishop, 2017; Jenson et al., 2020). Mu power suppression is also observed in music perception (Mercadié et al., 2014; Ross et al., 2022; Wang et al., 2023), and is further amplified when music perception is coupled with visual actions (Tanaka, 2021).

Recent studies using naturalistic continuous stimuli have begun investigating the neural tracking of continuous stimulus signals like speech envelope (Aller et al., 2022; Harding et al., 2019; Park et

al., 2016). Cerebral-acoustic coherence (CACoh) analysis allows the measurement of the coherence between neural signals and continuous audio envelope signals (Harding et al., 2019; Peelle et al., 2013; Teng et al., 2024). Therefore, both Mu power suppression with naturalistic continuous stimuli and CACoh analysis can enable us to investigate how visual actions may modulate Mu entrainment to music amplitude envelope, and the correlation between neural entrainment to music and music-induced pleasure, which has never been done before.

In the present study, we recorded simultaneous electrodermal activity (EDA) and EEG signals when participants who did not know how to play violin listened to naturalistic vocal or violin music under four conditions: audio-visual congruent (AVc, where music and visual performance were matched), audio-visual incongruent (AVic, where audio tracks were switched across videos), visual only (VO), and audio only (AO) (Figure 1). We aimed to investigate whether visual congruency influences music-induced pleasure and whether this effect is mediated by sensorimotor mechanisms.

We first analyzed subjective pleasant/liking ratings and EDA to assess the effect of visual congruency on musical pleasure using both subjective and objective indices. We then extracted Mu wave EEG activity (Mu power suppression and Mu entrainment to music signal) to investigate how visual congruency affects sensorimotor integration during music perception. Given our hypothesis that musical pleasure is modulated by visual congruency through sensorimotor integration, and considering that none of the participants could play the violin, we predicted that visual congruency would primarily influence the pleasure derived from vocal music, as participants could not utilize movement cues specific to playing the violin, such as fine finger movements but could interpret vocal movements, which are universally familiar. Lastly, we conducted correlation analyses and mediation analyses among visual congruency, musical pleasure, and Mu wave activity (power suppression and entrainment) to investigate their relationship during neural processing of music.

Results

Congruent visual information increases subjective pleasure ratings to music

After each music piece under AVc, AVic, and AO conditions, participants gave their liking, pleasure, arousal, and familiarity ratings to the music pieces. After each video under the VO condition, participants gave their liking ratings to the video.

One-way repeated measures analysis of variance (ANOVA) found that the main effects of visual conditions (AVc, AVic, and AO) on subjective pleasure ratings and liking ratings to music were significant only for vocal music (Pleasure: Vocal: $F(2, 30) = 10.09$, $p < .001$, $\eta_p^2 = 0.402$; Violin: $F(2, 30) = 1.19$, $p = 0.166$, $\eta_p^2 = 0.113$; Liking: Vocal: F(2, 30) = 6.97, p = 0.003, $\eta_p^2 = 0.32$; Violin: F(2, 30) = 2.23, p = 0.125, $\eta_p^2 = 0.13$). No significant main effect of visual conditions was found on arousal and familiarity ratings (Arousal: Vocal: $F(2, 30) = 0.35$, $p = 0.705$, $\eta_p^2 = 0.02$; Violin: $F(2, 30) = 1.93$, $p = 0.163$, $\eta_p^2 = 0.11$; Familiarity: Vocal: F(2, 30) = 0.02, p = 0.985, $\eta_p^2 = 0.00$; Violin: F(2, 30) = 0.71, p = 0.501, $\eta_p^2 = 0.05$). We also perform two-way repeated ANOVA (visual conditions × music type) to examine whether the effect of visual conditions on liking and pleasure ratings are different between vocal and violin music. However, no significant interaction effect was found (Pleasure: $F(2, 60) = 0.88$, $p = 0.403$, $\eta_p^2 = 0.03$; Liking: $F(2, 60) = 0.83$, $p = 0.829$, $\eta_p^2 = 0.01$).

As is shown in Figure 2a, post hoc analysis found that congruent visual information increased listeners' pleasure ratings and liking ratings to vocal music compared to AVic condition (pleasure rating: $t(30) = 3.87$, $p_{\text{corrected}} < .001$, Cohen's $d = 0.79$; liking rating: t(30) = 3.53, $p_{\text{corrected}} = 0.004$, Cohen's d = 0.79). And no significant effect was found for violin music (pleasure rating: $t(30) = 1.93$, $p_{\text{corrected}} = 0.191$, Cohen's $d = 0.28$; liking rating: t(30) = 2.09, $p_{\text{corrected}} = 0.137$, Cohen's d = 0.31).

To examine whether congruent visual information increased music-induced pleasure and liking ratings or incongruent visual information decreased music-induced pleasure and liking ratings, we compared pleasure and liking ratings between audiovisual conditions and AO condition (Wilcoxon signed rank tests were performed if the data violated the assumption of normality. Z-scores denote the statistics for Wilcoxon signed rank tests, and r values denote the effect size). Pleasure rating under AVc condition to vocal music was higher than that under AO condition ($Z = 2.82$, $p_{\text{corrected}} = 0.015$, $r = 0.51$), while liking rating did not increase under AO condition ($t(30) = 2.04$, $p_{\text{corrected}} = 0.151$, Cohen's $d = 0.42$). However, pleasure rating and liking rating were not significantly different between AVic and AO conditions (pleasure rating: $t(30) = 1.27$, $p_{\text{corrected}} = 0.642$, Cohen's $d = 0.64$; liking rating: AO: $Z = 2.00$, $p_{\text{corrected}} = 0.138$, $r = 0.36$). The results suggest that the impact of visual congruency on pleasure ratings to music can be attributed to the enhancement of these ratings by congruent visual information.

Congruent visual information increases skin conductance responses during music listening

SCR frequency and amplitude were extracted from EDA data to investigate the effect of visual congruency on electrodermal activity. As shown in Figure 2b, paired Wilcoxon signed rank test revealed that subjects showed significantly more SCRs under AVc condition than AVic condition when listening to vocal music ($Z = 1.98$, $p = 0.048$, $r = 0.38$), while no significant difference was found in violin music (Figure 2b). Subjects also showed marginally larger summed SCR amplitude when listening to vocal music with congruent visual movement compared to AVic condition ($Z = 1.87$, $p = 0.061$, $r = 0.36$), while no effect was found in violin music. No significant effect of music type on SCR difference between AVc and AVic conditions was found (SCR frequency: $Z = 1.28$, $p = 0.201$, $r = 0.25$; summed SCR amplitude: $Z = 1.40$, $p = 0.162$, $r = 0.27$).

Visual information inhibits Mu-band power independently of congruency

Mu-band power was extracted for each visual condition to investigate whether the presence of visual information inhibited Mu-band power and whether visual congruency modulated Mu-band power during music perception. As is shown in Figure 3a, we found that Mu-band power under conditions with visual information was significantly inhibited compared to AO condition at almost all frontal, temporal, and occipital electrodes (AVc vs. AO, AVic vs. AO, p_{fwe} < 0.05) for both vocal and violin music. However, we did not find any electrodes that showed a significant difference in Mu power between AVc and AVic conditions.

Additionally, Mu-band power was significantly suppressed at all central and frontal electrodes under AVc condition compared to AO condition during vocal music perception. However, for violin music, Mu-band power under AVc condition did not show significant suppression at Cz, C1, C2, and C3 electrodes, which typically showed Mu suppression when a person performed or observed an action (Hobson and Bishop, 2017), possibly because our participants do not know how to play the violin.

Visual congruency modulates neural entrainment to vocal music in the Mu band

In addition to Mu-band power, CACoh analysis was performed to investigate the effect of visual congruency on the neural entrainment to the music signal in the Mu band. As shown in Figure 3b, we found that CACoh vocal music was significantly greater at frontal electrodes (FP1, FP2, AF7, AFz, AF4, AF8, F5, FC3, and C5) under AVc condition than AVic condition (p_{fwe} < 0.05). However, no electrodes showed a significant effect of violin music CACoh.

Robust linear regression analysis was conducted to investigate the association of the music CACoh difference at electrodes that showed a significant visual congruency effect with the SCR difference, and with subjective rating difference between AVc and AVic conditions. We found that the average increased CACoh across significant channels could significantly predict the increased pleasure ratings in vocal music (Figure 3c, β = 0.39, p = 0.001). Also, increased CACoh could predict larger summed SCR amplitude for vocal music (β = 0.35, p = 0.034).

Greater Mu entrainment to violin music with congruent visual information is correlated to increased music-induced pleasure and Mu suppression

We did not find a visual congruency modulation effect to violin music-induced pleasure and neural entrainment generally. However, considering individual differences in audiovisual violin music perception, we hypothesized that subjects who showed greater Mu suppression or Mu entrainment to music under visual congruent condition compared to visual incongruent condition could demonstrate greater music-induced pleasure. Therefore, we performed robust linear regression analysis to examine whether Mu power difference across electrodes that showed significant suppression in audiovisual vocal music but not in audiovisual violin music (Cz, C1, C2, and C3) or average CACoh difference at electrodes that showed greater neural entrainment to vocal music under AVc condition (FP1, FP2, AF7, AFz, AF4, AF8, F5, FC3, and C5) could predict violin music-induced pleasure difference. And we also examine the relationship between the average CACoh difference and Mu power difference.

Robust linear regression results showed that greater Mu entrainment under AVc condition compared to AVic condition predicted greater summed SCR amplitude, SCR frequency, and Mu suppression for violin music (Figure 3c, summed SCR amplitude: β = 0.34, p = 0.005; SCR

frequency: $\beta = 0.43$, $p < 0.001$; Mu suppression: $\beta = -0.36$, $p < 0.001$).

Additionally, greater Mu suppression was correlated with increased subjective arousing ratings ($\beta = -0.16$, $p = 0.021$), and longer musical training years could marginally predict greater Mu suppression ($\beta = -0.15$, $p = 0.065$).

Increased neural entrainment to vocal music mediates increased music-induced pleasure rating with congruent visual information

Since we found that visual congruency modulated both Mu entrainment to music and music-induced pleasure to vocal music, we explored whether enhanced entrainment to music mediates increased music-induced pleasure under AVc condition. The electrodes that showed significantly greater Mu entrainment to vocal music were included in the within-subject mediation analysis. As shown in Figure 4a, we found a significant indirect effect of visual congruency on subjective pleasure ratings through increased mean music CACoh value across all significant electrodes (Figure 4a, visual congruency → average CACoh value → subjective pleasure rating, 95% CI, 0.01 to 0.23). Additionally, we found a full mediation effect of the mean Mu entrainment, since visual congruency was not correlated with pleasure rating after controlling the mean Mu entrainment (from $\beta = 0.27$, $p < 0.001$ to $\beta = 0.17$, $p > 0.05$). These results indicated that the Mu entrainment fully explained the visual modulation on subjective pleasure rating.

Increased Mu suppression mediates the association between increased neural entrainment to music and music-induced SCR change

In violin music, we discovered a significant correlation between the difference in Mu entrainment, SCR frequency, and the summed SCR amplitude difference under AVc and AVic conditions, as well as the difference in Mu suppression. We also conducted the mediation analysis to explore further relationships between neural and psychophysiological indices. As shown in Figure 4b, we found a significant full mediation effect of Mu suppression between neural entrainment to music and SCR frequency difference (Δ CACoh value → Δ Mu suppression → Δ SCR frequency, 95% CI, 0.02 to 0.25, from $\beta = 0.46$, $p = 0.019$ to $\beta = 0.17$, $p > 0.05$). These results suggest that increased Mu suppression fully explains the relationship between neural entrainment to music and the physiological difference. However, the mediation effect was not significant when neural entrainment was set as the mediator (Δ Mu suppression→ Δ CACoh value → Δ SCR frequency, 95% CI, -0.26 to 0.05). Also, we did not find a significant mediation effect when summed SCR amplitude difference was set as the dependent variable (Δ CACoh value → Δ Mu suppression → Δ summed SCR amplitude, 95% CI, -0.03 to 0.40; Δ Mu suppression→ Δ CACoh value → Δ summed SCR amplitude, 95% CI, -0.30 to 0.10).

Discussion

In this study, we present evidence that congruent performer movements enhanced listeners' musical rewards compared to incongruent movements, at both behavioral and physiological levels using naturalistic music pieces. Such enhancement is likely attributable to the facilitation of sensorimotor integration among listeners, as indicated by Mu entrainment to music during the listening experience. Notably, all participants were non-violin players, and these findings were exclusively observed in vocal music but not violin music. This further substantiates that enhanced sensorimotor integration serves as a crucial neural mechanism for augmenting musical rewards through visual movement information.

Several previous behavioral studies have found that music performance appreciation, as subjectively rated by listeners, increases when observers view congruent performers' movement (Vines et al., 2011, 2006; Tsay, 2013; Eaves et al., 2020; Griffiths and Reay, 2018; Platz and Kopiez, 2012). Consistent with this phenomenon, we found that listeners' subjective pleasure ratings were modulated by the congruency of visual information, with increased pleasure under congruent conditions. Furthermore, SCRs, which serve as reliable objective indicators of arousal and/or reward levels in response to music stimuli (Martínez-Molina et al., 2016; Mas-Herrero et al., 2014; Salimpoor et al., 2009; Solberg and Dibben, 2019; Zatorre, 2015), were found to increase in frequency and amplitude under congruent visual conditions. These results support the hypothesis that visual congruency between the performer's action and the music significantly heightens musical pleasure.

Previous studies have also found improved speech perception when observers simultaneously view congruent lip movements (Park et al., 2016; Zhang and Du, 2022; Zion Golumbic et al., 2013), which provides both content information and predictive timing of the speech (Peelle and Sommers, 2015). Several MEG and EEG studies have found better neural tracking of speech amplitude envelope when listening to speech with congruent lip movements (Crosse et al., 2016, 2015; Giordano et al., 2017; Park et al., 2016; Zion Golumbic et al., 2013). Analogous effects were observed in the present study regarding music envelope tracking through coherence analysis, suggesting that similar mechanisms, that is, performers' movements, may provide content information and predictive timing of the vocal music, may enhance neural entrainment to music.

The role of the motor system in music-induced pleasure remains underexplored. A recent meta-analysis failed to find a significant association between cortical motor areas and music-induced pleasure (Mas-Herrero et al., 2021). Previous fMRI studies have shown that the motor system is involved in rhythmic information processing (Gordon et al., 2018), and several studies have found correlations between cortical and subcortical motor areas and beat-related pleasure (Kornysheva et al., 2010; Matthews et al., 2020; Trost et al., 2014). In vocal music perception, visual vocal motor information provides music features like timing, pitch, and amplitude. Listeners could employ these visual cues to better predict the upcoming music, thereby increasing predictability and reducing uncertainty with congruent visual movements of the performers. Previous studies investigating how music surprisal and entropy interact with pleasure rating found that music with intermediate levels of complexity (surprisal) and uncertainty (entropy) received higher pleasure ratings (Cheung et al., 2019; Gold et al., 2019). Therefore, we speculate that clearer content and greater predictive information perceived through congruent visual information leads to a better internal representation and, consequently, greater pleasure for the listeners.

In addition, greater coherence between neural activity and the music envelope in the Mu band was only observed in vocal music, not in violin music, over sensorimotor and frontal electrodes. None of the participants in this study could play the violin, while all could sing to some extent. Mu-band

EEG signals reflect activities of the human mirror neuron system (Hobson and Bishop, 2017). Recent neuroimaging and neural modulation studies have found that the recruitment of the motor system is important for speech perception, especially under adverse conditions (Du et al., 2014; Hickok et al., 2011; Liang et al., 2023; Murakami et al., 2015; Zhang et al., 2023; Zhang and Du, 2022). Although not crucial, previous studies have found that motor system is also involved during music listening, especially when people have acquired knowledge of playing the musical instrument (Gordon et al., 2018; Herholz et al., 2016; Lahav et al., 2007; Srinivasan et al., 2020). Several musical training experiments have found that after training, listeners showed greater sensorimotor integration during music listening (Herholz et al., 2016; Herholz and Zatorre, 2012; Srinivasan et al., 2020; Wollman et al., 2018; Zatorre et al., 2007). These findings may explain why listeners benefited from vocal music production video. With knowledge of vocal music production, they can extract more information from congruent visual vocal movements. Conversely, since the participants did not know how to play the violin, they could not derive as much music information from specific violin-playing motions, rendering visual congruency less influential on violin music perception.

Although we did not find a significant modulation of visual congruency on violin music processing or pleasure, some subjects may still gather information from violin playing movements, such as bow movements that are correlated with note onset and sound volume. Robust regression analysis revealed a significant correlation between enhanced Mu entrainment and violin music-induced pleasure, indicating that listeners who exhibited greater Mu entrainment enhancement to violin music were more likely to experience greater violin music-induced pleasure with congruent visual movements. These results imply that sensorimotor integration plays an important role in visual enhancement of music-induced pleasure.

Mu band power is typically suppressed when observing or imitating actions of others (Hobson and Bishop, 2017). In this study, we found Mu suppression under both AVc and AVic conditions compared to the AO condition, with no difference between AVc and AVic conditions. Instead, the finding that visual congruency significantly modulated the coherence between Mu band neural activities and music signals provides a new approach to investigate the role of Mu-band brainwave in sensory perception.

There are several limitations in this study. First, due to the use of natural stimuli, the effect of performers' general movements, such as body sway and fine movements in violin playing, cannot be separated. To address this issue, future studies could employ motion capture to record audiovisual music stimuli, allowing separate investigation of different motion components. In addition, the relatively poor spatial resolution of EEG signals prevents us from identifying specific brain areas in the frontal lobe that contribute to the enhanced brain entrainment to music and music-induced pleasure. In future studies, neuroimaging techniques with higher spatial resolution like MEG and fMRI could be employed to obtain more accurate spatial information.

In sum, this study provides empirical evidence that music-induced pleasure can be modulated by observing performers' body movements, as reflected in both subjective ratings and skin conductance responses. Since performers' movements provide information about music production, we propose that sensorimotor integration of the audience is enhanced during music perception, leading to better predictions and reduced uncertainty, factors that are believed to influence pleasure (Zatorre, 2023). Enhanced sensorimotor integration is a significant factor through which an audience experiences greater music-induced pleasure with congruent visual movements compared to incongruent visual movements.

Method

Participants

The sample size was estimated using power analyses (α = 0.05, power = 0.8) in G*Power 3.1 (Faul et al., 2009). Using a two-tailed paired t-test, 33 participants would be sufficient to detect the effect size (Cohen's d = 0.51) estimated by a previous meta-analysis (Platz and Kopiez, 2012) with 80% power. Based on the power analysis and experimental design, 32 participants (18 females) were recruited for this study. One subject was removed from further analysis due to fatigue. Additionally, the first two subjects' EEG data were invalid due to a technical error and were discarded from further analysis, although their behavioral and EDA data were retained. Four subjects' EDA data were excluded from the analysis because they exhibited no skin conductance responses after taking a deep breath and holding it for several seconds, indicating non-responsiveness in terms of EDA (Braithwaite et al., 2013).

All the participants (18 females, 18 – 33 years old) were right-handed, non-Mandarin speakers, without any neurological or psychiatric disorders, and had no more than 2 years of training in string instruments or vocals (although some had training in other instruments). The Barcelona Musical Reward Questionnaire (BMRQ) scores of all participants were above 65 (indicating no music anhedonia) (Mas-Herrero et al., 2013). All participants had signed the written consent approved from the Institutional Review Board of McGill University.

Stimuli and procedure

Twenty naturalistic music pieces (10 vocal and 10 violin) were selected from YouTube and Bilibili. These music pieces were all Japanese or Chinese pop music, with vocal music lyrics in Mandarin. The pieces were selected so that they would likely be unfamiliar to our Western listeners who did not speak Mandarin. Consequently, any observed effects of congruency can be attributed solely to lower-level cues, such as the movements of the mouth, fingers and arms, and their synchronization with the musical sounds. The videos only contained the performer's movements against a static background, with no other musical accompaniment. Dynamic parts of the videos, such as rolling titles, were blurred by a Gaussian function to ensure they were invisible to the subjects. Music videos were trimmed to approximately 80-second segments. A pilot study was conducted to further refine the selection of music pieces, where 13 participants were asked to rate their liking of the music. Eight pieces were selected based on the liking ratings and the length of the music, and were used in the psychophysics and EEG experiment. The average sound intensity of each audio was normalized according to the root mean square value, and the sampling rate was 48 kHz. The video resolution was 1280 × 720 pixels, with a frame rate of 30 frame/s. The video and audio parts of the clips were combined according to 4 conditions (see Figure 1 for the experimental design).

Audiovisual congruent condition (AVc). Music and visual performance were matched (i.e. the intact video was used);

Audiovisual incongruent condition (AVic). Music and visual performance were mismatched by switching audio tracks across videos (see below);

Audio-only condition (AO). Only music was played with a static cross at the center of the screen;

Visual-only condition (VO); Only video was played without music.

Note that in the AVic condition, vocal video was combined only with vocal music, and violin video only with violin music. The same videos and music from the AVc condition were used to generate the AVic stimuli. Each pair of pieces was always combined with each other. For example, if vocal music 1

was combined with vocal video 2, vocal music 2 was combined with vocal video 1, and vice versa. Pairs of music pieces with similar durations were chosen to be combined with each other. If one peice was longer or shorter, the video was trimmed to match the music duration. Each trial began with a 3-second static image, which was the first frame of the video.

The experiment consisted of 8 blocks (violin/vocal × 4 conditions). Each vocal or violin block contained 4 trials with different conditions. After each trial, participants rated (0-9) their liking, pleasure, arousal, and familiarity with the music if music was presented, and their liking (0-9) of the visual performance if the visual performance was presented. Conditions were counterbalanced using a Latin square design. Music type was also balanced using an ABBA design (A denoted a violin-vocal block sequence, and B denoted a vocal-violin block sequence). Therefore, an ABBA sequence denoted violin/vocal/vocal/violin/vocal/violin/violin/vocal block). ABBA sequences were presented to half of the subjects, and BAAB sequences were presented to the other half.

Behavioral data analysis

Behavioral responses to vocal and violin music were analyzed separately. We performed a one-way repeated measures ANOVA to investigate the effect of visual conditions on ratings for vocal and violin music. Greenhouse–Geisser correction was performed if the sphericity assumption was violated. Post hoc analysis further examined the difference between specific conditions (AVc vs AO, AVic vs AO, AVc vs AVic). Wilcoxon signed rank tests were performed if the data violated the assumption of normality, as assessed by the Shapiro-Wilks normality test. Bonferroni correction was applied for multiple comparisons. Statistical analysis was conducted in R with the package bruceR (Bao, 2020) and visualized with ggplot2 (Wickham, 2009).

EDA data acquisition and analysis

EDA data were collected using the BIOPAC MP150 system with the EDA 100C amplifier module. Two electrodes with isotonic gel were placed on the volar surfaces of the distal phalanges of the right index and middle finger. The sampling rate was 2000 Hz, and the gain was set to 5 μS/V.

To keep the consistency of the data length of each condition, the first 77 seconds of each trial were analyzed. EDA was analyzed using Neurokit2 (Makowski et al., 2021), a python toolbox, and custom scripts. Default parameters were used. The number and summed amplitude of SCR in each condition were extracted. Note that, only SCRs that reached a peak between 2 seconds after the onset and the end of the music were considered, as the latency of SCR is about 2 seconds.

To account for baseline differences in SCR activity among subjects, the AO condition was treated as the baseline, and the relative difference in SCR indexes under AVc and AVic conditions entered the latter statistical analyses.

Two-sampled paired Wilcoxon signed rank tests were performed to compare SCR indexes between AVc and AVic conditions for vocal and violin music separately, as the differences violated the normal distribution assumption (Shapiro-Wilk normality test, both $p < 0.042$).

EEG data acquisition and processing

EEG data were acquired from 64 Ag/AgCl electrodes using Brain Products actiCAP active electrode system, including 2 electrooculogram (EOG) electrodes. Impedance of all electrodes was kept below 25 kΩ, sufficient for high-quality data collection according to the actiCAP system manual. All channels were referenced to the average of the left and right mastoid channels (M1, M2).

EEG data analysis was conducted using the Fieldtrip toolbox (Oostenveld et al., 2011) and

custom scripts. Similar to EDA data analysis, only the first 77 seconds of each trial were analyzed. Raw EEG data were first bandpass filtered from 1 to 40 Hz, and downsampled to 90 Hz. Independent component analysis (ICA) was performed to remove the artifacts from eye blinks, eye movements, and heartbeat. Mastoid channels and EOG channels were removed in the latter analysis.

Spectral analysis

To investigate how visual performance affects Mu wave activity, spectral analysis was performed at each electrode. Preprocessed data were segmented into non-overlapping 1-second intervals. Power spectrum was computed and averaged across each condition using the function "ft_freqanalysis" provided in FieldTrip. Frequencies of interest were from 1 to 40 Hz in steps of 0.5 Hz. Spectral power from 8 to 13 Hz was averaged to get the overall mu wave power under each condition.

Cerebral-acoustic coherence (CACoh) analysis

CACoh analysis calculated cross-spectrum coherence between neural signals and music amplitude envelopes, which allows us to quantify phase-locked neural responses to acoustic signals at specific frequencies (Harding et al., 2019; Peelle et al., 2013; Teng et al., 2024). Audio signals were filtered using a bank of fourth-order gammatone filters with 16 center frequencies logarithmically spanning from 100 Hz to 4000 Hz. Hilbert transformation extracted the amplitude envelope of each band. The average of the 16 amplitude envelopes was downsampled to 90 Hz and bandpass filtered from 1 to 40 Hz to match the EEG signal for CACoh analysis.

We employed the function 'mscohere' in MATLAB to calculate magnitude-squared coherence. Magnitude-squared coherence between EEG signals at each electrode and the amplitude envelope of the music was estimated from 1 to 40 Hz in steps of 0.5 Hz across each condition. Coherence estimate was averaged between 8 to 13 Hz to assess the neural entrainment of Mu waves to the music.

Cluster-based permutation test

The cluster-based permutation test in FielfTrip was performed to correct for multiple comparisons in spectral and CACoh analyses. Two-sided paired t-tests assessed differences in CACoh values between AVc and AVic conditions at each electrode. The sum of the T-statistics for statistically significant electrodes with a cluster was calculated as the cluster-level statistic. 5000 Monte Carlo-based permutations were conducted to obtain the permutation distribution. The cluster-level alpha was set to 0.05.

Regression and mediation analysis

Robust linear regression analysis was performed to investigate the correlation between neural indices and music-induced pleasure using the "robustbase" package in R (Maechler et al., 2023).

Within-participant mediation analysis was performed using the "JSmediation" package in R (Yzerbyt et al., 2018) to investigate whether music processing mediated the enhanced pleasure with valid visual movements. For vocal music, the average CACoh difference of significant electrodes between AVc and AVic conditions was set as the mediator variable. Visual congruency was the binary predictor, and pleasure rating was the dependent variable. For violin music, the average CACoh difference of the electrodes were the predictor, SCR frequency or summed amplitude difference was the dependent variable, and Mu suppression difference was the mediator. Note that,

the “JSmediation” package reported original regression coefficients instead of robust regression coefficients for mediation analysis. Confidence intervals of the within-participant indirect effect were computed using the Monte Carlo method with 5000 iterations. The alpha threshold was set to 0.05.

Acknowledgments

We thank Alberto Ara, Arielle Rabinowitz, Kazuma Mori, and Oscar Bedford for help with experiment setup.

Funding: This work was supported by a grant from the Canadian Institutes of Health Research to RJZ. RJZ is supported by the Canada Research Chair program and by the Fondation pour l’Audition (Paris).

Author contributions: Conceptualization: L.Z., and R.J.Z. Methodology: L.Z., and R.J.Z. Analysis: L.Z. Investigation: L.Z. Writing—original draft: L.Z. Writing—review and editing: L.Z., Y.D., and R.J.Z. Visualization: L.Z., and R.J.Z. Supervision: Y.D., and R.J.Z.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper. The behavioral, EDA, and EEG data that support the findings of this study are available on OSF.

Reference

Aller M, Økland HS, MacGregor LJ, Blank H, Davis MH. 2022. Differential Auditory and Visual Phase-Locking Are Observed during Audio-Visual Benefit and Silent Lip-Reading for Speech Perception. *J Neurosci* **42**:6108-6120. doi:10.1523/JNEUROSCI.2476-21.2022

Bao H-W-S. 2020. bruceR: BRoadly Useful Collections and Extensions of R functions.

Bonini L, Rotunno C, Arcuri E, Gallese V. 2022. Mirror neurons 30 years later: implications and applications. *Trends Cogn Sci* **26**:767-781. doi:10.1016/j.tics.2022.06.003

Braithwaite JJ, Watson DG, Jones R, Rowe M. 2013. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* **49**:1017-1034.

Cheung VKM, Harrison PMC, Meyer L, Pearce MT, Haynes J-D, Koelsch S. 2019. Uncertainty and Surprise Jointly Predict Musical Pleasure and Amygdala, Hippocampus, and Auditory Cortex Activity. *Curr Biol* **29**:4084-4092.e4. doi:10.1016/j.cub.2019.09.067

Crosse MJ, Butler JS, Lalor EC. 2015. Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *J Neurosci* **35**:14195-14204. doi:10.1523/JNEUROSCI.1829-15.2015

Crosse MJ, Di Liberto GM, Lalor EC. 2016. Eye can hear clearly now: Inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. *J Neurosci* **36**:9888-9895. doi:10.1523/JNEUROSCI.1396-16.2016

Dalla Bella S, Peretz I, Rousseau L, Gosselin N. 2001. A developmental study of the affective value of tempo and mode in music. *Cognition* **80**:B1-B10. doi:10.1016/S0010-0277(00)00136-0

Du Y, Buchsbaum BR, Grady CL, Alain C. 2014. Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc Natl Acad Sci U S A* **111**:7126-31. doi:10.1073/pnas.1318738111

Eaves DL, Griffiths N, Burridge E, McBain T, Butcher N. 2020. Seeing a drummer's performance modulates the subjective experience of groove while listening to popular music drum patterns. *Music Sci* **24**:475-493. doi:10.1177/1029864919825776

Faul F, Erdfelder E, Buchner A, Lang A-G. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods* **41**:1149-1160. doi:10.3758/BRM.41.4.1149

Fox NA, Bakermans-Kranenburg MJ, Yoo KH, Bowman LC, Cannon EN, Vanderwert RE, Ferrari PF, van IJzendoorn MH. 2016. Assessing Human Mirror Activity With EEG Mu Rhythm: A Meta-Analysis. *Psychol Bull* **142**:291-313. doi:10.1037/bul0000031

Gagnon L, Peretz I. 2003. Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cogn Emot* **17**:25-40. doi:10.1080/02699930302279

Giordano BL, Ince RAA, Gross J, Schyns PG, Panzeri S, Kayser C. 2017. Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *eLife* **6**:e24763. doi:10.7554/eLife.24763

Gold BP, Pearce MT, Mas-Herrero E, Dagher A, Zatorre RJ. 2019. Predictability and Uncertainty in the Pleasure of Music: A Reward for Learning? *J Neurosci* **39**:9397-9409. doi:10.1523/JNEUROSCI.0428-19.2019

Gomez P, Danuser B. 2007. Relationships between musical structure and psychophysiological measures of emotion. *Emotion* **7**:377-387. doi:10.1037/1528-3542.7.2.377

Gordon CL, Cobb PR, Balasubramaniam R. 2018. Recruitment of the motor system during music listening: An ALE meta-analysis of fMRI data. *PLOS ONE* **13**:e0207213. doi:10.1371/journal.pone.0207213

Griffiths NK, Reay JL. 2018. The Relative Importance of Aural and Visual Information in the Evaluation of Western Canon Music Performance by Musicians and Nonmusicians. *Music Percept* **35**:364-375. doi:10.1525/mp.2018.35.3.364

Harding EE, Sammler D, Henry MJ, Large EW, Kotz SA. 2019. Cortical tracking of rhythm in music and speech. *NeuroImage* **185**:96-101. doi:10.1016/j.neuroimage.2018.10.037

Herholz SC, Coffey EBJ, Pantev C, Zatorre RJ. 2016. Dissociation of Neural Networks for Predisposition and for Training-Related Plasticity in Auditory-Motor Learning. *Cereb Cortex* **26**:3125-3134. doi:10.1093/cercor/bhv138

Herholz SC, Zatorre RJ. 2012. Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron* **76**:486-502. doi:10.1016/j.neuron.2012.10.011

Hickok G, Houde J, Rong F. 2011. Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron* **69**:407-422. doi:10.1016/j.neuron.2011.01.019

Hobson HM, Bishop DVM. 2017. The interpretation of mu suppression as an index of mirror neuron activity: past, present and future. *R Soc Open Sci* **4**:160662. doi:10.1098/rsos.160662

Jenson D, Bowers AL, Hudock D, Saltuklaroglu T. 2020. The Application of EEG Mu Rhythm Measures to Neurophysiological Research in Stuttering. *Front Hum Neurosci* **13**:458. doi:10.3389/fnhum.2019.00458

Kornysheva K, von Cramon DY, Jacobsen T, Schubotz RI. 2010. Tuning-in to the beat: Aesthetic appreciation of musical rhythms correlates with a premotor activity boost. *Hum Brain Mapp* **31**:48-64. doi:10.1002/hbm.20844

Lahav A, Saltzman E, Schlaug G. 2007. Action Representation of Sound: Audiomotor Recognition Network While Listening to Newly Acquired Actions. *J Neurosci* **27**:308-314. doi:10.1523/JNEUROSCI.4822-06.2007

Liang B, Li Y, Zhao W, Du Y. 2023. Bilateral human laryngeal motor cortex in perceptual decision of lexical tone and voicing of consonant. *Nat Commun* **14**:4710. doi:10.1038/s41467-023-40445-0

Maechler M, Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, Koller M, Conceicao ELT, Palma MA di. 2023. robustbase: Basic Robust Statistics.

Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behav Res Methods* **53**:1689-1696. doi:10.3758/s13428-020-01516-y

Martínez-Molina N, Mas-Herrero E, Rodríguez-Fornells A, Zatorre RJ, Marco-Pallarés J. 2016. Neural correlates of specific musical anhedonia. *Proc Natl Acad Sci* **113**:E7337-E7345. doi:10.1073/pnas.1611211113

Mas-Herrero E, Maini L, Sescousse G, Zatorre RJ. 2021. Common and distinct neural correlates of music and food-induced pleasure: A coordinate-based meta-analysis of neuroimaging studies. *Neurosci Biobehav Rev* **123**:61-71. doi:10.1016/j.neubiorev.2020.12.008

Mas-Herrero E, Marco-Pallares J, Lorenzo-Seva U, Zatorre RJ, Rodriguez-Fornells A. 2013. Individual Differences in Music Reward Experiences. *Music Percept* **31**:118-138. doi:10.1525/mp.2013.31.2.118

Mas-Herrero E, Zatorre RJ, Rodriguez-Fornells A, Marco-Pallarés J. 2014. Dissociation between musical and monetary reward responses in specific musical anhedonia. *Curr Biol* **24**:699-704.

Matthews TE, Witek MAG, Lund T, Vuust P, Penhune VB. 2020. The sensation of groove engages motor and reward networks. *NeuroImage* **214**:116768. doi:10.1016/j.neuroimage.2020.116768

Mercadié L, Caballe J, Aucouturier J-J, Bigand E. 2014. Effect of synchronized or desynchronized

music listening during osteopathic treatment: An EEG study. *Psychophysiology* **51**:52-59. doi:10.1111/psyp.12143

Michalak J, Troje NF, Fischer J, Vollmar P, Heidenreich T, Schulte D. 2009. Embodiment of Sadness and Depression—Gait Patterns Associated With Dysphoric Mood. *Psychosom Med* **71**:580. doi:10.1097/PSY.0b013e3181a2515c

Murakami T, Kell CA, Restle J, Ugawa Y, Ziemann U. 2015. Left Dorsal Speech Stream Components and Their Contribution to Phonological Processing. *J Neurosci* **35**:1411-1422. doi:10.1523/JNEUROSCI.0246-14.2015

Oostenveld R, Fries P, Maris E, Schoffelen J-M. 2011. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput Intell Neurosci* **2011**:1-9. doi:10.1155/2011/156869

Overy K, Molnar-Szakacs I. 2009. Being together in time: Musical experience and the mirror neuron system. *Music Percept* **26**:489-504. doi:10.1525/mp.2009.26.5.489

Park H, Kayser C, Thut G, Gross J. 2016. Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife* **5**:e14521. doi:10.7554/eLife.14521

Peelle JE, Gross J, Davis MH. 2013. Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cereb Cortex* **23**:1378-1387. doi:10.1093/cercor/bhs118

Peelle JE, Sommers MS. 2015. Prediction and constraint in audiovisual speech perception. *Cortex* **68**:169-181. doi:10.1016/j.cortex.2015.03.006

Platz F, Kopiez R. 2012. When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Percept* **30**:71-83. doi:10.1525/mp.2012.30.1.71

Ross JM, Comstock DC, Iversen JR, Makeig S, Balasubramaniam R. 2022. Cortical mu rhythms during action and passive music listening. *J Neurophysiol* **127**:213-224. doi:10.1152/jn.00346.2021

Salimpoor VN, Benovoy M, Longo G, Cooperstock JR, Zatorre RJ. 2009. The rewarding aspects of music listening are related to degree of emotional arousal. *PloS One* **4**:e7487.

Sievers B, Polansky L, Casey M, Wheatley T. 2013. Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc Natl Acad Sci* **110**:70-75. doi:10.1073/pnas.1209023110

Solberg RT, Dibben N. 2019. Peak Experiences with Electronic Dance MusicSubjective Experiences, Physiological Responses, and Musical Characteristics of the Break Routine. *Music Percept* **36**:371-389. doi:10.1525/mp.2019.36.4.371

Srinivasan N, Bishop J, Yekovich R, Rosenfield DB, Helekar SA. 2020. Differential Activation and Functional Plasticity of Multimodal Areas Associated with Acquired Musical Skill. *Neuroscience* **446**:294-303. doi:10.1016/j.neuroscience.2020.08.013

Tanaka S. 2021. Mirror Neuron Activity During Audiovisual Appreciation of Opera Performance. *Front Psychol* **11**.

Teng X, Larrouy-Maestri P, Poeppel D. 2024. Segmenting and Predicting Musical Phrase Structure Exploits Neural Gain Modulation and Phase Precession. *J Neurosci* **44**. doi:10.1523/JNEUROSCI.1331-23.2024

Trost W, Frühholz S, Schön D, Labbé C, Pichon S, Grandjean D, Vuilleumier P. 2014. Getting the beat: entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage* **103**:55-64. doi:10.1016/j.neuroimage.2014.09.009

Tsay CJ. 2013. Sight over sound in the judgment of music performance. *Proc Natl Acad Sci U S A*

110:14580-14585. doi:10.1073/pnas.1221454110

Vines BW, Krumhansl CL, Wanderley MM, Dalca IM, Levitin DJ. 2011. Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition* **118**:157-170. doi:10.1016/j.cognition.2010.11.010

Vines BW, Krumhansl CL, Wanderley MM, Levitin DJ. 2006. Cross-modal interactions in the perception of musical performance. *Cognition* **101**:80-113. doi:10.1016/j.cognition.2005.09.003

Wang Y, Siu TSC, Cheung H. 2023. Effect of music emotion on mu and beta oscillations. *Psychol Music* **51**:1489-1500. doi:10.1177/03057356221145960

Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wollman I, Penhune V, Segado M, Carpentier T, Zatorre RJ. 2018. Neural network retuning and neural predictors of learning success associated with cello training. *Proc Natl Acad Sci* **115**:E6056-E6064. doi:10.1073/pnas.1721414115

Yzerbyt V, Muller D, Batailler C, Judd CM. 2018. New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *J Pers Soc Psychol* **115**:929-943. doi:10.1037/pspa0000132

Zatorre R. 2023. From Perception to Pleasure: The Neuroscience of Music and Why We Love It. Oxford, New York: Oxford University Press.

Zatorre RJ. 2015. Musical pleasure and reward: mechanisms and dysfunction. *Ann N Y Acad Sci* **1337**:202-211. doi:10.1111/nyas.12677

Zatorre RJ, Chen JL, Penhune VB. 2007. When the brain plays music: auditory-motor interactions in music perception and production. *Nat Rev Neurosci* **8**:547-558. doi:10.1038/nrn2152

Zhang L, Du Y. 2022. Lip movements enhance speech representations and effective connectivity in auditory dorsal stream. *NeuroImage* **257**:119311. doi:10.1016/j.neuroimage.2022.119311

Zhang L, Wang X, Alain C, Du Y. 2023. Successful aging of musicians: Preservation of sensorimotor regions aids audiovisual speech-in-noise perception. *Sci Adv* **9**:eadg7056. doi:10.1126/sciadv.adg7056

Zion Golumbic E, Cogan GB, Schroeder CE, Poeppel D. 2013. Visual input enhances selective speech envelope tracking in auditory cortex at a "Cocktail Party." *J Neurosci* **33**:1417-1426. doi:10.1523/JNEUROSCI.3675-12.2013

Figures

Figure 1. Experiment conditions. A denotes audio music stimuli and V denotes visual movements stimuli. AV congruent condition: Music stimuli and visual body movements stimuli were from the same video. AV incongruent condition: Music stimuli and visual body movements stimuli were from different video. Audio only condition: Only music stimuli was presented and a static cross was presented on the screen. Visual only condition: Only silent visual body movements movement video was presented.

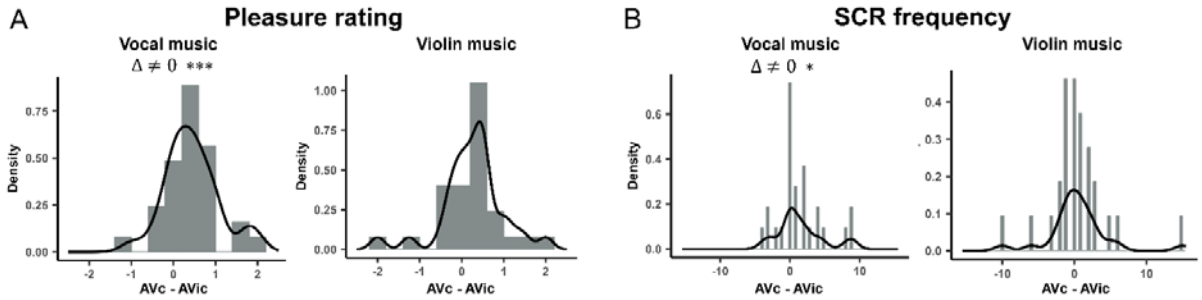

Figure 2. Histogram and density plot of the difference of subjective pleasure ratings and SCR frequency. Subjective pleasure ratings (A) and relative SCR frequency (B) were higher under AVc condition than AVic condition for vocal music but not for violin music.

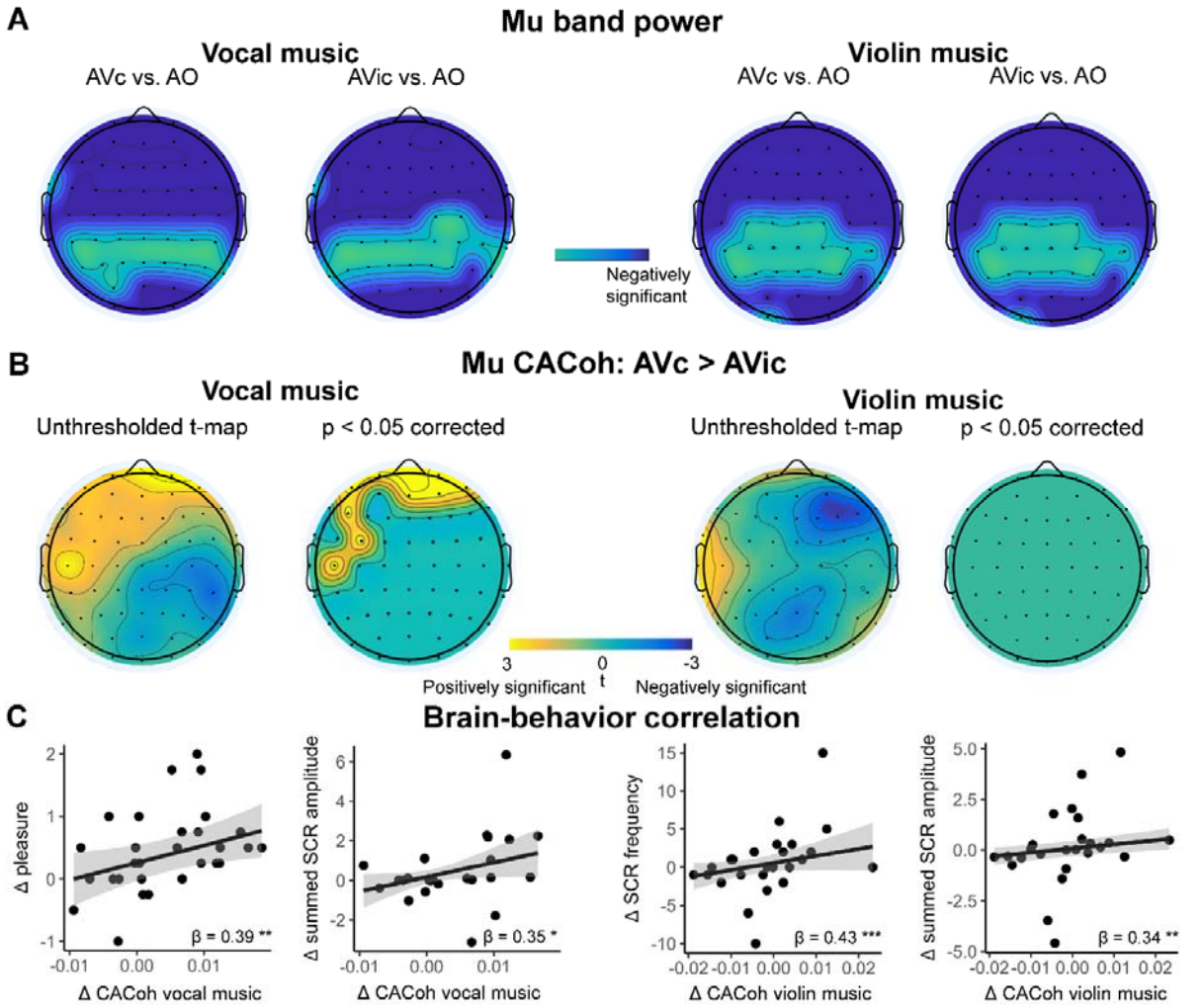

Figure 3. (A) Mu-band power was suppressed under AVc and AVic conditions compared to AO condition for both vocal and violin music. (B) Mu entrainment to music, as calculated by cerebral-acoustic coherence (CACoh) between Mu-band signal and music amplitude envelope, was enhanced under AVc condition compared to AVic condition only for vocal music. (C) Enhanced neural entrainment to music was correlated to greater subjective pleasure ratings and SCR indices for both vocal and violin music.

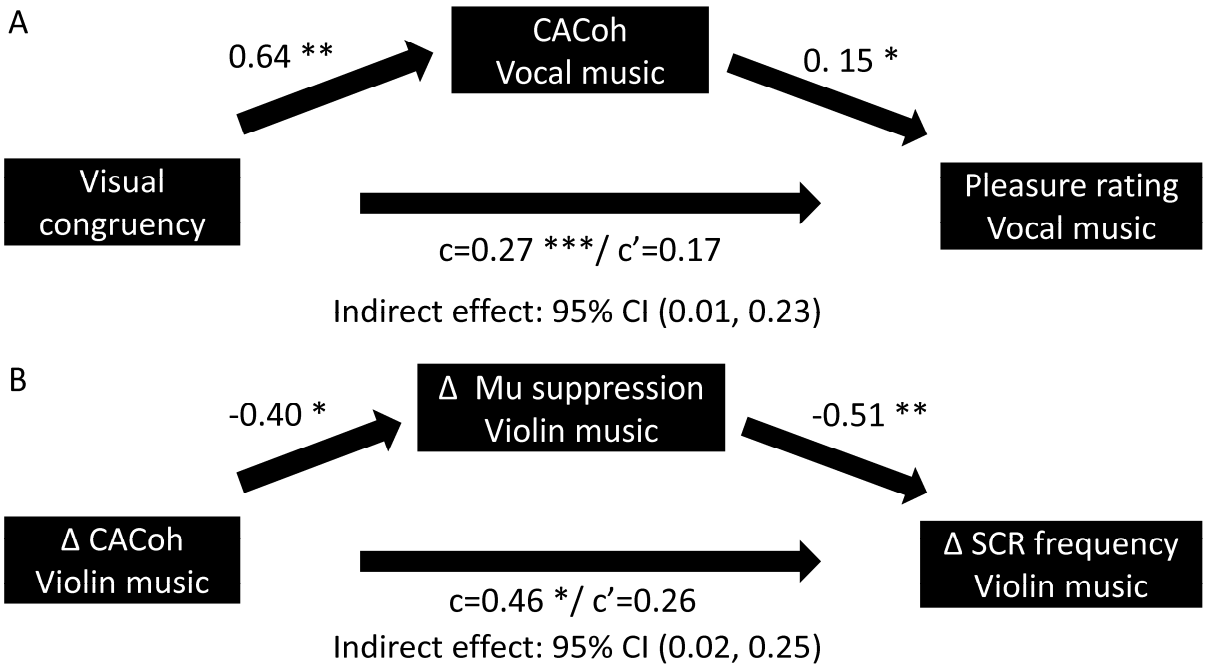

Figure 4. Within-subject mediation analysis results. (A) Mu entrainment to vocal music mediated the visual modulation on subjective pleasure ratings. (B) Enhanced Mu power suppression to violin music under AVc condition was the significant mediator of the correlation between enhanced Mu entrainment and greater SCR frequency to violin music under AVc condition compared to AVic condition. CACoh: Cerebral-acoustic coherence.