Explicit processing of melodic structure in congenital amusia can be improved by redescription-associate learning

Article

Accepted Version

Jiang, J., Liu, F. ORCID: https://orcid.org/0000-0002-7776-0222, Zhou, L., Chen, L. and Jiang, C. (2023) Explicit processing of melodic structure in congenital amusia can be improved by redescription-associate learning. Neuropsychologia, 182. 108521. ISSN 1873-3514 doi: https://doi.org/10.1016/j.neuropsychologia.2023.108521 Available at https://centaur.reading.ac.uk/110788/

To link to this article DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2023.108521

Publisher: Elsevier

Explicit Processing of Melodic Structure in Congenital Amusia can be Improved by Redescription-Associate Learning

Jun Jiang[1], Fang Liu[2], Linshu Zhou[1], Liaoliao Chen[3], Cunmei Jiang[1*]

[1] Music College, Shanghai Normal University, Shanghai, China, 200234

[2] School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK, RG6 6AL

[3] Foreign Languages College, Shanghai Normal University, Shanghai, China, 200234

* Corresponding author at: Music College, Shanghai Normal University, 100 Guilin Road, Xuhui District, Shanghai 200234, China.

E-mail addresses: cunmeijiang@shnu.edu.cn (C. Jiang)

Highlights

- Amusia is characterized by intact implicit and impaired explicit pitch perception.
- We developed a redescription-associate learning method to treat amusics' deficits.
- Prior to training, amusics failed to explicitly process melodic structure.
- After training, amusics showed normal performance at behavioral and neural levels.
- Training effects on melodic processing in amusia maintained at 3-month follow-up.

Abstract

Congenital amusia is a neurodevelopmental disorder of musical processing. Previous research demonstrates that although explicit musical processing is impaired in congenital amusia, implicit musical processing can be intact. However, little is known about whether implicit knowledge could improve explicit musical processing in individuals with congenital amusia. To this end, we developed a training method utilizing redescription-associate learning, aiming at transferring implicit representations of perceptual states into explicit forms through verbal description and then establishing the associations between the perceptual states reported and responses via feedback, to investigate whether explicit processing of melodic structure could be improved in individuals with congenital amusia. Sixteen amusics and 11 controls rated the degree of expectedness of melodies during EEG recording before and after training. In the interim, half of the amusics received nine training sessions on melodic structure, while the other half received no training. Results, based on effect size estimation, showed that at pretest, amusics but not controls failed to explicitly distinguish regular from irregular melodies and to exhibit an early right anterior negativity (ERAN) in response to irregular endings. At posttest, trained but not untrained amusics performed as well as controls at both the behavioral and neural levels. At the 3-month follow-up, the training effects still maintained. These findings present novel electrophysiological evidence of neural plasticity in the amusic brain, suggesting that redescription-associate learning may be an effective method to remediate impaired explicit processes for individuals with other neurodevelopmental disorders who have intact implicit knowledge.

Keywords: explicit processing, redescription-associate learning, pitch structure, congenital amusia, ERAN

1 Introduction

Congenital amusia (amusia hereafter) is a neurodevelopmental disorder of musical pitch processing (Ayotte et al., 2002; Peretz et al., 2002) with prevalence of 1.5%–4% among the general population (Henry & McAuley, 2010; Kalmus & Fry, 1980; Nan et al., 2010; Peretz & Vuvan, 2017). Amusics show impaired explicit processing of pitch, including fine-grained pitch detection (e.g., Hyde & Peretz, 2004; Jiang et al., 2011; Liu et al., 2010; Whiteford & Oxenham, 2017) and pitch direction perception (e.g., Foxton et al., 2004; Jiang et al., 2013; Liu et al., 2012; Loui et al., 2008; Lu et al., 2017), which have also been linked to a lack of a P3b, an ERP component related to conscious detection of small pitch deviances (Lu et al., 2016; Moreau et al., 2013; Peretz et al., 2005), and pitch short-term memory (e.g., Albouy et al., 2013; Albouy et al., 2019; Graves et al., 2019; Tillmann, Lévêque, et al., 2016). Unlike explicit performance, the amusic brain can implicitly detect small pitch changes (Mignault Goulet et al., 2012; Moreau et al., 2009, 2013; Quiroga-Martinez et al., 2021), by showing normal early ERP components such as the mismatch negativity. The deficits in explicit processing of musical pitch may be due to impaired pitch awareness in amusia (Loui, 2016; Peretz, 2016; Tillmann et al., 2015).

Apart from low-level processes such as pitch detection/discrimination, the dissociation between implicit and explicit pitch perception has also been observed in higher-level processes including pitch structure processing in amusia. Specifically, amusics demonstrate preserved implicit knowledge of melodic (Lévêque et al., 2022; Omigie et al., 2012; Tillmann et al., 2014; Tillmann, Lalitte, et al., 2016; Weiss & Peretz, 2022) and harmonic (Tillmann et al., 2012) syntactic structures, being able to process melodic structure implicitly by evoking an early right anterior negativity (ERAN; Zendel et al., 2015), an index of the processing of musical pitch structure violations (Koelsch et al., 2000;

Sun et al., 2020). However, intact knowledge of musical pitch structure does not facilitate amusics' pitch structure processing in an explicit manner. This is because they are impaired in explicitly processing melodic (Jiang et al., 2016; Omigie et al., 2012; Tillmann, Lévêque, et al., 2016) and harmonic structures (Jiang et al., 2016), as manifested by an absence of the ERAN (Peretz et al., 2009; Zhou et al., 2019) and late integrative components such as the N5 and P600/LPC (Peretz et al., 2009; Zendel et al., 2015; Zhou et al., 2019). Furthermore, it is worth noting that impaired explicit higher-level pitch structure processing (e.g., rating how well the notes or chords in a sequence followed one another in an expected manner) is uncorrelated with their low-level pitch detection/discrimination (Jiang et al., 2016; Omigie et al., 2012; Tillmann, Lalitte, et al., 2016; Zhou et al., 2019). Similarly, previous studies (Jiang et al., 2016; Tillmann, Lalitte, et al., 2016) have also reported amusics' pitch structure processing is uncorrelated with pitch discrimination (same vs. different) on the Montreal Battery of Evaluation of Amusia (MBEA), a diagnostic tool for amusia (Peretz et al., 2003; Vuvan et al., 2018). These findings indicate that the processing of musical pitch structure may be independent of pitch discrimination in amusia.

A big challenge in research on amusia has been the remediation of the musical disorder. Prior research has mainly focused on treating low-level pitch perception deficits (Anderson et al., 2012; Liu et al., 2017; Mignault Goulet et al., 2012; Whiteford & Oxenham, 2018; Wilbiks et al., 2016). Some studies show that deficits in pitch perception in amusia cannot be altered through broad-brush music training methods, e.g., daily song listening over 4 weeks (Mignault Goulet et al., 2012), group singing over 7 weeks (Anderson et al., 2012), or vocal training over 18 months (Wilbiks et al., 2016). These null results may be attributed to the fact that neither song listening nor singing training is well suited to remediate amusics' pitch perceptual deficits. However, other approaches using training tasks similar to the test tasks show that amusics' sensitivity to pitch change can be improved through a 2-week pitch direction training (Liu et al., 2017) and a 4-day pitch discrimination training (Whiteford & Oxenham, 2018). These findings suggest that amusics can be responsive to pitch training when the training paradigms are specific and targeted.

Unlike low-level pitch perception, higher-level musical pitch structure processing is at the core of intramusical meaning understanding (Budd, 1995; Gruhn, 2005; Kertz-Welzel, 2005), which concerns the processing of tonal syntax that organizes hierarchically discrete pitch events (tones, intervals, and chords) into melodic or harmonic sequences (Marmel et al., 2011; Patel, 2003). However, no study has explored the effects of music training on amusics' deficits in explicit processing of pitch structure. Given that amusic individuals have implicit tonal knowledge, but fail to consciously access the knowledge (e.g., Omigie et al., 2012; Peretz et al., 2009; Tillmann, Lalitte, et al., 2016; Zendel et al., 2015), it would be worthwhile to examine how to make the stored implicit knowledge externalized, and then to see if the externalization of implicit knowledge can facilitate explicit processing of musical pitch structure.

In this scenario, we developed a training method employing redescription-associate learning for amusics. The core of this method is to externalize implicit knowledge through verbal reports, and then to establish the associations between the reported perceptual states and the responses through feedback. A basic premise of this method is that a stimulus can be mentally represented in many manners (Cermeño-Aínsa, 2021; Paivio, 1990; Pearson & Kosslyn, 2015; Quilty-Dunn, 2020). Among these representations, one perceptual state may be easier to verbalize than the others. For example, a mental representation of a melody may be a sequence of notes, an image of a (pitch) upward/downward movement, and/or a certain feeling induced by the melody in the mind. For some listeners, the feeling induced by the melody may be easily expressed, while others may consider it easier to describe the image of an upward/downward movement.

The redescription-associate learning method consists of two steps. *The first step* aims to teach amusic participants to transfer implicit representations or knowledge into explicit forms. Specifically, after making judgements, amusic participants would be asked to report verbally their perceptual states pertaining to music listening (e.g., emotions and imagery). This is because verbal report is considered to play a key role in representational change (Cleeremans, 2019; Karmiloff-Smith, 1986, 1992, 1994; Nelson, 1996) and could drive an individual to recode implicit representations of perceptual states into explicit representations, according to representational redescription model (Karmiloff-Smith, 1986, 1992, 1994). Indeed, it has been suggested that implicit representations can be converted into explicit forms through a verbalization training (Park, 2013, 2015; Park & Choi, 2006; Park et al., 2008) or group discussion (Pine & Messer, 1998). For example, Pine and Messer (1998) investigated whether children who were at the implicit level on a balance beam task at pretest would benefit from group discussion with other children and progress to more explicit levels at posttest. During the training session, the experimental group was offered opportunities for discussion or verbal interaction with other children about whether the beam could be balanced before completing each beam; the control group was only asked to perform the balance beam task without any discussion. The results showed that children in the experimental group (40%) were more likely to progress to the explicit level at posttest than those in the control group (0%). These findings suggest that group discussion can facilitate the conversion of children’s implicit knowledge into explicit knowledge. Likewise, verbalization training can also drive representational change in number conservation (Park et al., 2008) and drawing (Park, 2013, 2015). *The second step* of redescription-associate learning aims to teach amusic participants to establish the associations between the perceptual states they reported and the responses through feedback and then to consolidate memory of such associations by continuing training. This is because associate learning can establish a connection between two unrelated elements (Olson & Ramirez, 2020; Pavlov, 1955; Skinner, 1953). Although being unaware of the associations, amusic participants could build the connection between a reported perceptual state and a certain response through receiving and learning from feedback, given that feedback has been viewed as playing a critical role in the formation of associations (Bischoff-Grethe et al., 2009; Butler et al., 2013; Kuklick & Lindner, 2021; Marsh et al., 2012; Skinner, 1958). For example, Kuklick and Lindner (2021) examined the association between questions and responses through verification feedback. In the training phase, participants answered multiple-choice questions that were related to scientific concepts. The experimental group received right/wrong feedback after each question, while the control group received no feedback. Results showed that the experimental group had greater improvement in recall performance than the control group. The benefit of feedback also emerged in the research on reading comprehension (Butler et al., 2013), general knowledge retrieval (Marsh et al., 2012), and mathematics problem solving (Brown & Alibali, 2018). Considering that memorization increases the strength of associations (Naveh-Benjamin et al., 2007), amusic participants would then be required to memorize these associations and be provided with more training trials to further consolidate the associations. By doing so, participants would produce the appropriate responses when encountering similar stimuli again during testing.

Following the steps mentioned above, we investigated whether explicit processing of melodic structure can be improved in individuals with amusia through the redescription-associate learning. First, given that music structure is typically represented in cadence (i.e., the end or closure of a phrase), we manipulated the regularities of melodic sequences to investigate the processing of melodic structure, with regular melodies ending on the tonic and irregular melodies ending on the supertonic. Second, in order to evaluate the efficacy of music training, we randomly divided the amusics into two groups: trained and untrained. Three months posttest, trained amusics completed a follow-up test to examine the maintenance of the training effect. Finally, because the detection of melodic structure violations is typically associated with the ERAN elicited by syntactic irregularities (Brattico et al., 2006; Koelsch, 2012; Peretz et al., 2009; Zendel et al., 2015), we included an EEG study focusing on the ERAN effect. We hypothesized that the amusic participants would explicitly distinguish regular from irregular melodies and elicit a similar ERAN effect to controls after completing the training, but not before the training.

2 Method

2.1 Participants

Sixteen amusic and 11 typically developing undergraduate and postgraduate students who spoke Mandarin Chinese as their native language and had not received any formal music training participated in this study. The musical abilities of these participants were assessed by the MBEA that consists of six subtests—Scale, Contour, Interval, Rhythm, Meter, and Memory (Peretz et al., 2003; Vuvan et al., 2018). The first three pitch-based subtests require participants to discriminate between different melodies, by detecting an out-of-key note, an altered contour or interval, respectively. Participants were diagnosed as amusics if they scored 65 or below on the melodic composite score (sum of the scores on the Scale, Contour, and Interval subtests) (Liu et al., 2010) and below 78% correct on the MBEA global score (Peretz et al., 2003). The amusics were randomly divided into two groups: The trained group underwent the training program, whereas the untrained group received no training. A two-alternative forced choice AXB paradigm was used to measure pitch perception thresholds (Jiang et al., 2013). Table 1 shows the participants' characteristics. As can be seen, the three groups were matched in age, sex, education, and pitch change detection threshold, but there were significant between-groups differences in all MBEA scores and pitch direction discrimination threshold. Post hoc pairwise comparisons (*n* = 3) using the Games–Howell procedure (Sauder & DeMars, 2019) indicated that no significant difference in performance was observed between the two amusic groups, while manifesting worse performance than the control group (see Supplementary Table S1).

Insert Table 1, about here.

All participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). All had normal hearing and reported to have normal or corrected-to-normal vision. None reported history of neurological or psychiatric diseases. Ethical approval was granted by the Human Ethics Committee of Shanghai Normal University, and all participants gave written informed consent and were paid for their participation.

2.2 Stimuli

The stimuli for this study included 216 melodic pairs that were composed according to the rules of the Western tonal system. For each melody, there were five tones ranging from G_3 (196 Hz) to A_4 (440 Hz), where the first four tones had a duration of 0.5 s and the final tone lasted 1 s. Each melody began with the tonic, the mediant, or the dominant, while the fourth tone was either the subdominant or the leading tone. In keeping with previous studies on music structure (Jiang et al., 2016; Koelsch, 2012; Sun et al., 2020; Zhou et al., 2019), we manipulated the music-syntactic regularities at the end of these melodies. Thus, each pair was given a regular (the tonic, C_4) and an irregular ending (the supertonic, D_4) which violated music-syntactic regularities in Western tonal music (see Figure 1). Owing to the constraints of the first, fourth, and final tones, the final tone (the tonic in the regular or the supertonic in the irregular melodies) in each melody might be repeated once at most. The two melodies of a pair were identical in melodic contour, and overall, the pitch distances between the fourth and final tones in the regular and irregular conditions were equal (6 semitones). Specifically, in one half of the melody pairs, the pitch distance (1 semitone) between the last two tones of the regular melodies was smaller than that (3 semitones) of the irregular melodies, while in the other half of the melody pairs, the pitch distance (5 semitones) between the last two tones of the regular melodies was larger than that (3 semitones) of the irregular melodies.

Insert Figure 1, about here.

All pairs of melodies were then randomly divided into the testing (72 pairs) and the training stimuli (144 pairs). For the testing stimuli, 72 regular melodies ended on the tonic, while the 72 irregular melodies ended on the supertonic. To examine the effect of sensory novelty (Koelsch et al., 2007; Tillmann et al., 2019) on musical structural processing, we computed the ratio of occurrence of the tonic and the supertonic in the regular and irregular melodies. The data were not normally distributed (Shapiro-Wilk test $W = 0.64$, $p < .001$)

and we thus employed JASP (Version 0.17.1) to conduct a Wilcoxon signed-rank test. The result revealed that the ratio of the supertonic (*Mdn* = .20, interquartile range [*IQR*] = .20) in the irregular melodies was higher than that of the tonic (*Mdn* = .20, *IQR* = .20) in the corresponding regular melodies ($T = 126.00$, $z = 2.34$, $p = .012$, $r_{rb} = .647$, 95% CI [.225, .865]). This result indicates that the supertonic may sound more salient than the tonic, as pitch salience is associated with frequency of occurrence of tones: the higher the frequency of occurrence, the more salient the tones (Krumhansl, 1990; Krumhansl & Kessler, 1982; Lantz et al., 2020).

To model acoustic information stored in auditory short-term memory, we used the eaR (Version 0.2.1) Package in R (Version 4.0.3) to calculate tonal contextuality—correlations between the local and global pitch images, using method II. This measure reflects both the degree of similarity between the immediate pitch and its pitch context (Leman, 2000; Marmel et al., 2010) and the tension of a local pitch image with respect to the global pitch image (Bigand et al., 2014), with higher tonal contextuality values associated with higher degrees of similarity and lower tension levels between a local pitch and its global context (Bigand et al., 2014). Following previous work (Collins et al., 2014; Sears et al., 2019), the echo of the local image was kept at the default of 0.1 s, whereas the echo of the global image was set to 4 s to be compatible with the duration of echoic memory (Darwin et al., 1972). The data were analyzed with Wilcoxon signed-rank test, due to deviation from normality ($W = 0.85$, $p < .001$). Results showed that tonal contextuality values of irregular melodies (*Mdn* = .72, *IQR* = .06) were comparable to those of regular melodies (*Mdn* = .71, *IQR* = .07), $T = 1382.00$, $z = 0.38$, $p = .705$, $r_{rb} = .052$, 95%

CI [−.211, .307]. This result indicates that irregular melodies did not differ significantly from regular melodies in tension.

The training sessions comprised 144 melodic pairs. Each pair consisted of a regular melody ending on the tonic and an irregular melody ending on the supertonic, resulting in 288 melodies. These melodies were divided into nine blocks of 32 trials, 16 with regular endings and 16 with irregular endings. Like the testing stimuli, we calculated the ratio of the tonic and the supertonic. Because of nonnormally distributed ($W = 0.63$, $p < .001$), the Wilcoxon signed-rank test was used. We found that the ratio of the supertonic (*Mdn* = .20, *IQR* = .20) in the irregular melodies was higher than that of the tonic (*Mdn* = .20, *IQR* = .00) in the corresponding regular melodies ($T = 1176.00$, $z = 6.03$, $p < .001$, $r_{rb} = 1.000$, 95% CI [1.000, 1.000]). Likewise, we also calculated the tonal contextuality values. The data violated the normality assumption ($W = 0.86$, $p < .001$), and the Wilcoxon signed-rank test were thus performed. We found that irregular melodies (*Mdn* = .73, *IQR* = .06) had larger values than regular melodies (*Mdn* = .71, *IQR* = .04), $T = 7049.00$, $z = 3.65$, $p < .001$, $r_{rb} = .350$, 95% CI [.176, .504]. Together, these results suggest that irregular melodies may be more salient and less tense than regular melodies in the training stimuli.

All melodies were composed in C major and generated with Sibelius (Version 7.5) and played with a piano sound (Yamaha S90ES) using Cubase (Version 5.1). Sound files were recorded with a sampling rate of 44100 Hz, 16-bit resolution, and 705-kbps bit rate.

2.3 Procedure

Figure 2 illustrates the timeline of this study. As can be seen, the study included five stages. In stage 1 (pretest), all groups finished the melodic structure test with behavioral and EEG measures. After listening to a melody via Edifier R1200T loudspeakers (Edifier Technology Co., Ltd., Beijing, China), participants were required to rate how well the melody fitted with their expectation on a 7-point Likert item (1 = *very unexpected*, 7 = *very expected*) by pressing one of the number keys 1–7 on the computer keyboard. They were encouraged to use the whole range of the item. All melodies were presented pseudorandomly such that melodies with the same type of endings (regular or irregular) did not occur more than three times consecutively. Prior to the test, participants were provided with four practice trials to familiarize them with the task and were asked to adjust the sound volume to their most comfortable listening level.

Insert Figure 2, about here.

In stage 2 (training), only the training group received 9 sessions of training over 4.5 weeks. These training sessions were administered twice per week; each lasted about 40 min. During training, participants sat alone in front of a computer in the lab, listened to one melody at a time, and judged after the completion of music whether the tones of the melody followed one another in an expected manner by pressing the "F" or "J" key on a computer keyboard. The text "Correct" or "Incorrect" and the proportions of correct responses were then displayed on the computer screen. After making 1-3 consecutive correct judgements, participants were required to describe their perceptual experiences during listening to the melodies and to memorize the association between a certain perceptual state and a response (expected or unexpected). Because of a lack of music expertise, participants often utilized similes to describe their perceptual experiences. For example, an expected/regular melody was frequently described as "*This melody is like a complete sentence, a flat road, or a relaxed heart*.", whereas an unexpected/irregular melody was frequently described as "*This* *melody is like an incomplete sentence, a downhill/uphill road, or an anxious heart*." When an incorrect judgement was made, the participants were also encouraged to verbalize their perceptual experiences of the melody. They then listened to this melody and made a judgement again. If the participants still made an incorrect response, they would be presented with the regular and irregular versions of this melody in succession, and required to compare their perceptual experiences about the two versions. The duration of the training sessions was determined by participants' performance, for which the criteria for termination gradually increased over the nine sessions. Training ended if the amusic participants achieved at least 75% correct for the first two sessions, 80% correct for the middle five sessions, and 90% correct for the last two sessions. However, the criteria were not met by one participant during the Sessions 1, 2, 4, 5, and 7 and two participants during the Sessions 3, 6, 8 and 9. These training sessions were terminated after 40 min for these participants. For Sessions 1 and 2, participants undertook at least three blocks; for Sessions 3-7, participants learned through at least two blocks; for Sessions 8 and 9, participants went through at least one block.

In stages 3 (posttest) and 4 (3-month follow-up), the procedure was identical to that during pretest. All of the participants completed the measures, whereas only trained amusics were invited to complete the follow-up test 3 months after training considering that cognitive representations of basic music-syntactic regularities in adults are remarkably stable and are less influenced by short-term musical experience (Carrión & Bly, 2008; Koelsch & Jentschke, 2008). In stage 5 (10-month follow-up), both trained and untrained amusics were asked to return to the lab and complete the MBEA and pitch threshold tests to see if there was a transfer effect.

2.4 EEG Recording and Preprocessing

EEG activity was continuously acquired from 64 Ag/AgCl electrodes positioned on an elastic cap according to the international 10–20 system, using a NeuroScan Acquire 4.3 (Compumedics NeuroScan Inc., Charlotte, NC, USA). To monitor the eye movement artefact, vertical electrooculogram signals were recorded from two bipolar electrodes placed over the upper and lower eyelids of the left eye, while horizontal electrooculogram signals were recorded from two electrodes placed 1 cm lateral to the external canthi. The left mastoid electrode (M1) was used as reference, and the forehead electrode (GND) served as ground. EEG signals were sampled at 500 Hz, online filtered between 0.05 and 100 Hz, and amplified with an AC-coupled NeuroScan Synamps amplifier. Electrode impedances were kept below 5 kΩ.

The acquired EEG signals were preprocessed offline using the NeuroScan Edit software (Version 4.5). First, the raw EEG data were rereferenced to the mean between both mastoids. Next, ocular artifacts were removed using a regression procedure implemented in the NeuroScan software (Semlitsch et al., 1986). Continuous data were then filtered with a zero-phase shift 0.1–30 Hz band pass filter (24 dB/oct slope), and epoched for 0.2 s before and until 1 s after the final tone onset. After this, baseline correction (−0.2–0 s) was applied to the data. Subsequently, trials with artifacts exceeding ± 75 μV at any channel were automatically rejected. Finally, trials were averaged by each condition for each participant at each electrode.

2.5 Statistical Analyses

The data were first tested for normality with the Shapiro-Wilk test and/or homoscedasticity with Levene’s test using jamovi (Version 2.3.24), or sphericity with Mauchly’s test using IBM SPSS Statistics (Version 29), given that violations of the assumptions can result in high probability of type I error, relatively poor power, or inaccurate confidence intervals (Erceg-Hurn & Mirosevich, 2008; Wilcox, 2022; Wilcox & Serang, 2017). If the assumptions of analysis of variance (ANOVA) in factorial designs were met, then parametric tests were performed; if any assumption was violated, then nonparametric tests were conducted (Field, 2017). The aligned rank transform (ART) is a nonparametric procedure that can be used to examine both main and interaction effects in ANOVA (Elkin et al., 2021; Wobbrock et al., 2011). Briefly, the ART procedure first aligns the data separately for each effect. The aligned data are then ranked and a classical parametric ANOVA is performed on the aligned ranks for every effect (Feys, 2016; Wobbrock et al., 2011; Wobbrock & Kay, 2016). When there was a significant main or interaction effect, nonparametric multiple pairwise comparisons with the ART-C algorithm were followed up (Elkin et al., 2021).

The parametric ANOVAs were based on Type III sums of squares and implemented with jamovi. The nonparametric ANOVAs were also based on Type III sums of squares and run in the ARTool (Version 0.11.1) package in R (Version 4.2.2) and RStudio (Version 2022.07.2+576), where a linear mixed-effects model with subjects as a random effect was fit for every dependent variable using the art() function. Following an interaction or a main effect, post hoc pairwise comparisons with Holm-Bonferroni correction (Streiner, 2015) and no correction were conducted across all possible pairs. However, in the interest of space, we only reported the most interesting and meaningful comparisons.

For behavioral data, a nonparametric three-way ANOVA was conducted using a linear mixed-effects model with group (trained amusics, untrained amusics, controls), time (pretest, posttest) and regularity (regular, irregular) as fixed effects and subjects as a random effect, because the data partly violated normality or homoscedasticity. For ERP data, mean amplitude values in the time window of 150–250 ms after the final tone onset were first computed for two regions of interest (ROIs): left anterior (AF3, F1, F3, F5, FC1, FC3, FC5) and right anterior (AF4, F2, F4, F6, FC2, FC4, FC6). The time window was selected a priori on the basis of previous research on the ERAN that used the explicit task with a similar age group and more female participants (Fiveash et al., 2018). The appropriateness of this time window for the ERAN was confirmed by visual inspection of the current data. The ERAN is mainly generated at the anterior electrode sites and located primarily in the frontal regions (Koelsch, 2012), especially when there are more females than males in the sample (Koelsch et al., 2003). Hence, we focused on the average of the frontal electrodes, thereby reducing the familywise error rate (Kappenman & Luck, 2016; Luck, 2014; Luck & Gaspelin, 2017). The choice of frontal electrodes was also based on and was identical to prior studies that used 64 scalp electrodes (Koelsch et al., 2013; Zhou et al., 2019). Then, a parametric four-way ANOVA was conducted with group as a between-subjects factor and time, regularity, and hemisphere (left, right) as within-subjects factors, because all data were normally distributed (Ws ≥ 0.83, ps $\geq .061$) and had equal variances (Fs ≤ 2.46, ps $\geq .107$).

The p value from the conventional null hypothesis significance testing is the probability of observing a test statistic as extreme or more extreme than that observed when the null hypothesis is true, but it cannot measure the size of an effect or the importance of a result (e.g., Wasserstein & Lazar, 2016; Wasserstein et al., 2019). In contrast, effect sizes and confidence intervals (CIs) provide estimates of the magnitude of the effect and the precision of their estimates, and are often used directly to infer significance levels (e.g., Calin-Jageman & Cumming, 2019a, 2019b; Cohen, 1990; Kline, 2013; Nakagawa & Cuthill, 2007). Indeed, the American Psychological Association (APA, 2010, 2020), as well as some researchers (e.g., Cumming, 2013, 2014; Griffiths & Needleman, 2019;

Karadaghy et al., 2017), strongly recommend reporting effect sizes and their CIs and using them to interpret results and draw conclusions whenever possible. Previous studies have focused on effect sizes and CIs to infer a significant effect, regardless of whether *p* values are reported (e.g., Biderman et al., 2019; Birmingham et al., 2015; Brown-Iannuzzi et al., 2015; Caplan et al., 2019; Kubit1 & Janata, 2022; Samson et al., 2019; Saunders et al., 2018). If the CI for the effect size includes zero, the effect does not exist; if the CI excludes zero, the effect does exist (Cheung et al., 2022; Perdices, 2018; Schober et al., 2018; Sohn, 1982; Steiger, 2004). Furthermore, it is recommended that effect sizes of Cohen's $d = 0.20$, 0.50, and 0.80 (Cohen, 1988) and $\omega^2 = .01$, .06, and .14 (Kirk, 1996) should be used as minimum cutoffs to interpret small, medium, and large effects, respectively. In the present study, we employed effect sizes and CIs for interpreting the results. We used the effectsize package (Version 0.8.1) to calculate an effect size and its CI (d and ω_p^2 with 95% CI). The raw data and code are available in the Supplementary Material.

3 Results

3.1 Training

Trials were discarded if the reaction time was less than 200 ms (anticipatory responses) or longer than 3 *SD* above everyone's mean for each training session. Figure 3 illustrates the percentage of correct responses across the nine training sessions for trained amusics. Although the percentages did not depart from normality (Ws ≥ 0.85, ps ≥ .090), Mauchly's test was not available as the sample size was smaller than the number of repeated measurements. Therefore, the Greenhouse-Geisser correction ($\varepsilon = 0.57$) was applied to adjust the degrees of freedom (Barcikowski & Robey, 1984) from one-way repeated measures ANOVA with training session as a within-subjects variable.

Insert Figure 3, about here.

The ANOVA revealed a main effect of training session with a large effect size, $F(4.57, 32.01) = 6.39$, $p < .001$, $\omega_p^2 = .396$, 95% CI [.074, .569]. We performed all 36 post hoc pairwise comparisons between the different training sessions using paired sample *t* tests (see Supplementary Table S2), but only presented eight pairs of the comparisons here in the main text. This is because we were primarily interested in the differences in performance between the first training session and the later sessions. Although the *p* values for most comparisons became nonsignificant after Holm adjustment, their effect sizes were large (Cohen, 1988), indicating that amusics scored higher on Session 2 ($t(7) = 2.81$, $p_{uncorrected} = .026$, $p_{holm} = .736$, $d = 1.06$, 95% CI [0.12, 1.96]), Session 3 ($t(7) = 2.76$, $p_{uncorrected} = .028$, $p_{holm} = .739$, $d = 1.04$, 95% CI [0.11, 1.93]), Session 4 ($t(7) = 3.96$, $p_{uncorrected} = .005$, $p_{holm} = .180$, $d = 1.50$, 95% CI [0.41, 2.54]), Session 5 ($t(7) = 4.00$, $p_{uncorrected} = .005$, $p_{holm} = .176$, $d = 1.51$, 95% CI [0.42, 2.56]), Session 6 ($t(7) = 2.78$, $p_{uncorrected} = .027$, $p_{holm} = .739$, $d = 1.05$, 95% CI [0.11, 1.94]), Session 7 ($t(7) = 3.19$, $p_{uncorrected} = .015$, $p_{holm} = .442$, $d = 1.21$, 95% CI [0.22, 2.15]), Session 8 ($t(7) = 6.08$, $p_{uncorrected} < .001$, $p_{holm} = .018$, $d = 2.30$, 95% CI [0.89, 3.67]) and Session 9 ($t(7) = 4.96$, $p_{uncorrected} = .002$, $p_{holm} = .057$, $d = 1.87$, 95% CI [0.64, 3.07]) than Session 1. These results indicate that the training improved the performance increasingly.

3.2 Pretest versus Posttest

3.2.1 Behavioral Results

Amusics' and controls' ratings on the regular and irregular melodies are plotted in Figure 4. As can be seen, relative to the control group, the two amusic groups had difficulty in distinguishing regular from irregular melodies at pretest. However, trained but not untrained amusics differentiated regular from irregular melodies as controls at posttest, showing a positive effect of music training. These findings were confirmed by the three-way ANOVA on the aligned ranks, which revealed a significant effect of regularity with a large effect size (F(1, 72) = 151.27, $p < .001$, $\omega_p^2 = .670$, 95% CI [.545, .754]), but no effects of time (F(1, 72) = 5.55, $p = .021$, $\omega_p^2 = .058$, 95% CI [.000, .188]) and group (F(2, 24) = 3.87, $p = .035$, $\omega_p^2 = .175$, 95% CI [.000, .416]) despite significant p values. There were medium-to-large effect sizes for the two-way interactions of Time × Group (F(2, 72) = 5.54, $p = .006$, $\omega_p^2 = .108$, 95% CI [.004, .245]), Time × Regularity (F(1, 72) = 16.65, $p < .001$, $\omega_p^2 = .175$, 95% CI [.044, .330]), and Group × Regularity (F(2, 72) = 24.47, $p < .001$, $\omega_p^2 = .385$, 95% CI [.208, .523]). There was also a large effect size for the three-way interaction of Time × Group × Regularity (F(2, 72) = 12.64, $p < .001$, $\omega_p^2 = .237$, 95% CI [.077, .386]). All 66 pairwise comparisons (see Supplementary Table S3) were carried out post hoc, but only the six comparisons of interest were reported here, examining whether each group of participants was able to distinguish between regular and irregular melodies at pretest and posttest. Results showed that there were no significant differences in expectedness ratings between regular and irregular melodies for trained (t(72) = 1.89, $p_{uncorrected} = .063$, $p_{holm} = 1.000$, $d = 0.22$, 95% CI [−0.01, 0.46]) and untrained (t(72) = 1.74, $p_{uncorrected} = .085$, $p_{holm} = 1.000$, $d = 0.21$, 95% CI [−0.03, 0.44]) amusics at pretest, while controls gave higher ratings for regular than irregular melodies (t(72) = 8.39, $p_{uncorrected}$

< .001, $p_{\text{holm}} < .001$, $d = 0.99$, 95% CI [0.70, 1.27]). Nevertheless, unlike untrained amusics ($t(72) = 1.62$, $p_{\text{uncorrected}} = .110$, $p_{\text{holm}} = 1.000$, $d = 0.19$, 95% CI [−0.04, 0.42]), trained amusics ($t(72) = 8.18$, $p_{\text{uncorrected}} < .001$, $p_{\text{holm}} < .001$, $d = 0.96$, 95% CI [0.68, 1.24]) and controls ($t(72) = 8.61$, $p_{\text{uncorrected}} < .001$, $p_{\text{holm}} < .001$, $d = 1.01$, 95% CI [0.73, 1.30]) gave higher ratings for regular than irregular melodies at posttest. According to Cohen's (1988) criteria, effect sizes (*d*) of 0.96 for trained amusics and 1.01 for controls were considered large and comparable.

Insert Figure 4, about here.

3.2.2 ERP Results

The ERP results were in line with the behavioral results. Figure 5 and Figure 6 show the brain electric responses to regular and irregular melodies. As can be seen, no ERAN (150–250 ms) was evoked in trained or untrained amusics at pretest, while controls exhibited the ERAN with a bilaterally distributed topography over the scalp. However, like controls, trained but not untrained amusics showed a bilateral ERAN effect at posttest. These observations are verified by a large effect size for the interaction of Time × Group × Regularity (see Table 2). All 66 possible comparisons (see Supplementary Table S4) were done post hoc, but only the six comparisons of interest were presented here, examining whether each group of participants evoked an ERAN effect at pretest and posttest. Although a nonsignificant *p* value after Holm correction was observed, there was an ERAN effect with a medium effect size for controls ($t(24) = 3.39$, $p_{\text{uncorrected}} = .002$, $p_{\text{holm}} = .149$, $d = 0.69$, 95% CI [0.24, 1.13]) but not for either trained ($t(24) = 0.26$, $p_{\text{uncorrected}} = .799$, $p_{\text{holm}} = 1.000$,

$d = 0.05$, 95% CI [−0.35, 0.45]) or untrained ($t(24) = 0.32$, $p_{\text{uncorrected}} = .753$, $p_{\text{holm}} = 1.000$, $d = 0.07$, 95% CI [−0.34, 0.47]) amusics at pretest. Similarly, the ERAN effect with a medium-to-large effect size was found in controls ($t(24) = 3.17$, $p_{\text{uncorrected}} = .004$, $p_{\text{holm}} = .245$, $d = 0.65$, 95% CI [0.20, 1.08]) and trained amusics ($t(24) = 4.49$, $p_{\text{uncorrected}} < .001$, $p_{\text{holm}} = .010$, $d = 0.92$, 95% CI [0.43, 1.39]) but not in untrained amusics ($t(24) = −1.08$, $p_{\text{uncorrected}} = .289$, $p_{\text{holm}} = 1.000$, $d = −0.22$, 95% CI [−0.62, 0.19]) at posttest.

Insert Figure 5, about here.

Insert Figure 6, about here.

Insert Table 2, about here.

3.3 Posttest versus Follow-Up (Trained Amusics Only)

3.3.1 Behavioral Results

Figure 7 displays expectedness ratings of the regular and irregular melodies by seven of the eight trained amusics at posttest and at 3-month follow-up (one trained amusic participant dropped out of the study at follow-up). The ratings for all conditions were normally distributed (Ws ≥ 0.83, ps ≥ .082). Thus, the data were submitted to a parametric repeated measures ANOVA with time (posttest, follow-up) and regularity as the within-subjects factors. We found a significant main effect of regularity ($F(1, 6) = 21.22$, p =.004, ω_p^2 = .717, 95% CI [.121, .883]), with regular melodies (M = 5.48, SE = 0.29) receiving higher expectedness ratings than irregular melodies (M = 2.74, SE = 0.41). The main effect of time ($F(1, 6) = 0.28$, p = .614, ω_p^2 = −.099, 95% CI [.000, .000]) and the interaction of time and regularity ($F(1, 6) = 0.002$, p = .970, ω_p^2 = −.143, 95% CI [.000, .000]) were not significant, indicating that the training effect remained after 3 months.

Insert Figure 7, about here.

3.3.2 ERP Results

This training effect was also observed in the ERAN brain response (150–250 ms) as illustrated in Figure 8. Because the ERP amplitudes for all conditions were normally distributed (Ws ≥ 0.87, ps ≥ .177), we performed a parametric repeated measures ANOVA with time, regularity and hemisphere as the within-subjects factors. The ANOVA (see Table 3) revealed a large effect size for the main effect of regularity, because irregular endings elicited a more negative-going deflection (M = 3.91 μV, SE = 0.54) than regular endings (M = 5.49 μV, SE = 0.57). Furthermore, neither a main effect of time nor time-related interactions were found, indicating that the ERAN effect maintained from posttest to 3-month follow-up.

Insert Figure 8, about here.

Insert Table 3, about here.

3.4 Pretest versus Follow-Up for MBEA and Pitch Threshold Performance (Trained and Untrained Amusics)

Table 4 presents all MBEA scores and pitch perception thresholds of the seven trained and six untrained (two participants withdrew from the study at follow-up) amusics at pretest and 10-month follow-up. The two-way mixed ANOVAs with time (pretest, follow-up) as a within-subjects factor and group (trained amusics, untrained amusics) as a between-subjects factor were performed on the global MBEA score and the six subtest scores, respectively. There was a large-sized effect of time on the MBEA global score. This improvement was primarily related to great improvements in the Memory and Rhythm subtests. These results indicated that although there was no improvement in pitch-related subtests (Scale, Contour, and Interval), both trained and untrained amusics significantly improved their rhythm and memory performance from pretest to follow-up. In addition, no training effect was observed in any pitch threshold.

Insert Table 4, about here.

In order to investigate whether the training would influence controls' performance, four additional control participants were trained using the same protocol for the amusics. Individuals' performance during the nine training sessions demonstrated no obvious improvement (see Supplementary Figure S1). This led to no improvement in the explicit processing of melodic structure at both behavioral and neural levels after training. These findings suggest that controls have reached the best performance on our melodic structure task before training.

4 Discussion

Using EEG, we examined whether the redescription-associate learning method could improve explicit processing of melodic structure in individuals with amusia. Our *first* main finding is that through the redescription-associate learning method, impaired explicit processing of melodic structure in amusia was altered to be at a normal level. Specifically, prior to training, amusic individuals failed to consciously distinguish regular from irregular melodies. This finding is compatible with prior behavioral data (Jiang et al., 2016; Omigie et al., 2012; Peretz et al., 2009; Zendel et al., 2015) and ERP data (Peretz et al., 2009; Zendel et al., 2015) suggesting the deficits in processing melodic structure, as evidenced by the poor judgment and the absence of the ERAN evoked by irregular melodies in individuals with amusia. Such anomalies may result from gray (Mandell et al., 2007) and white (Albouy et al., 2013; Loui et al., 2009; Wang et al., 2017) matter abnormalities in the frontotemporal network (Loui, 2016; Peretz, 2016).

At posttest and the 3-month follow-up, however, trained but not untrained amusics performed as well as controls at both the behavioral and neural levels, as evidenced by their improved rating scores and the emergence of an ERAN effect, suggesting that the deficit in processing melodic structure can be ameliorated by the redescription-associate learning method. Indeed, previous work has indicated that low-level pitch discrimination (Whiteford & Oxenham, 2018) and pitch direction identification (Liu et al., 2017) in amusics can be improved with pitch perception training. Given that the dissociation between implicit and explicit processes was also found in pitch discrimination (e.g., Lu et al., 2017; Moreau et al., 2009; Moreau et al., 2013; Whiteford & Oxenham, 2017), redescription-associate learning may have the potential to improve amusics' lower-level pitch discrimination if training focused on explicit pitch discrimination. Furthermore, the observed behavioral and ERP effects indicate that trained amusics were able to process the hierarchy of stability of tones in tonality. Given that the ability to process the hierarchy of scale tones is the basis for the processing of tonal melodic and harmonic structures (Krumhansl & Keil, 1982; Patel, 2008), it is expected that trained amusics would be capable of processing melodies ending with other scale tones and harmonic sequences. In addition, apart from psychoacoustic attributes, tonal hierarchical information also carries emotional connotations of music (Jiang et al., 2017). Indeed, perceived tonal stability is correlated with perceived and felt musical tension (Bigand et al., 1996; Lehne et al., 2013; Steinbeis et al., 2006) and emotion (Maimon et al., 2022; Steinbeis et al., 2006). In this case, we may reasonably suppose that the reduced sensitivity to musical tension in amusia (Jiang et al., 2017) would be enhanced by the increased perception of tonal stability through redescription-associate learning. Such a hypothesis must, however, be empirically tested. Taken together, the present study extends the findings of previous training studies by showing that the processing of higher-level musical pitch structure in amusics can be improved with the redescription-associate learning method at both behavioral and neural levels, and indicates that an appropriate training method has a substantial impact on neural and behavioral plasticity.

Regarding what amusics actually learned during the training program, one may wonder whether it was indeed the learning of tonal structures or any other skills. Based on the current results, it can be inferred that what amusics learned during the training session is related to the tonal structures. First, the observed effects cannot be ascribed to the learning of melodic contour, because the regular and irregular melodies shared the same melodic contour. The learning was also unlikely to be about interval size, as the overall pitch distances between the fourth and final tones in the regular and irregular conditions were equal, with the first four tones being the same. Second, the observed effects cannot be ascribed to tone repetition or sensory influences. In our stimuli, the irregular ending tone had a higher ratio of occurrence and similar or higher tonal contextuality values than the regular ending tone, suggesting that the irregular ending tone was more salient and not more tense than the regular one. Based on these sensory features alone, it is impossible for the irregular ending to elicit a larger ERAN than the regular ending, as the ERAN reflects the processing of violations of tonal structures (Koelsch, 2012; Zendel et al., 2015), rather than that of tone repetition or sensory features. Finally, the learning was unlikely related to the possibility that trained amusics had simply learnt the difference between the tones C_4 and D_4, by memorizing the associations of these tones with their own responses after training. If this were the case, however, amusics would not have shown an ERAN response to the violations of tonal structures as demonstrated in our posttest and 3-month follow-up results. Thus, it can be concluded that amusics learned the tonal structures during the training program, rather than the other possibilities discussed above.

At 10-month follow-up, amusics' global MBEA scores increased as compared to pretest, because of their improved performance on the Rhythm and Memory subtests, but not on the pitch-related subtests. These findings are consistent with a previous study suggesting that trained amusics and controls (who received white noise localization training for 4 days) and untrained controls improved on the melodic and global MBEA performance (Whiteford & Oxenham, 2018). These improvements in our study most likely resulted from the retest or practice effect rather than the training effect. This is because both trained and untrained amusics showed improvement. Specifically, the improvement in memory appears to be related to instructions to participants. In the Memory subtest, participants need to decide whether they had heard the melody in the previous subtests (Peretz et al., 2003; Vuvan et al., 2018). When taking the MBEA for the first time, participants were not informed in advance that the melodies would be tested later and thus might not consciously remember them. When taking the test again, participants knew they would have a memory test later and thus might consciously remember them, which may lead to a better performance. Previous studies have also reported that memory performance would be enhanced by retest in children (Peretz et al., 2013) and adults (Lima et al., 2018). Likewise, the improvement in rhythm is possibly due to repeated exposure to the test materials. Nonetheless, this speculation needs to be further verified.

It is not surprising that our training on melodic structure did not have positive effects on pitch discrimination subtests of the MBEA and pitch detection/discrimination sensitivity. This is because the aims of the present study were to investigate the effects of music training on the processing of melodic syntactic structure, and the experimental design thus did not focus on low-level pitch perception. Therefore, the lack of the effect of music training on the MBEA performance may be attributed to the dissociation between low-level pitch perception and higher-level structural processing. Taking pitch dimension as an example, while pitch discrimination depends on sensitivity to pitch change, the processing of melodic structure is based on sensitivity to tonality, as sensitivity to tonality is a key premise for the processing of musical pitch structure (Krumhansl & Keil, 1982; Patel, 2008). Indeed, previous studies have also shown that impaired musical pitch structure processing is unrelated with performance on the Contour and Interval subtests of MBEA in amusia (Jiang et al., 2016; Tillmann, Lalitte, et al., 2016) and pitch perception thresholds

(Jiang et al., 2016; Omigie et al., 2012; Tillmann, Lalitte, et al., 2016; Zhou et al., 2019). Therefore, the present study confirms that low-level pitch perception and higher-level structural processing may be independent from each other.

Our *second* main finding is that the redescription-associate learning could be an effective method to remediate impaired explicit processes for individuals with amusia. Previous studies have demonstrated that individuals with amusia show impaired processing of musical structure in an explicit (e.g., Jiang et al., 2016; Peretz et al., 2009; Zhou et al., 2019), but not in an implicit manner (e.g., Omigie et al., 2012; Tillmann et al., 2012; Zendel et al., 2015). In this case, we designed the redescription-associate learning paradigm and demonstrated that impaired explicit processing of melodic structure could be remedied by this paradigm, suggesting neural and behavioral plasticity in amusics. Our findings confirm the role of representational reorganization through verbal reports in the externalization process (e.g., Cleeremans, 2019; Karmiloff-Smith, 1986; Park, 2015) and provide important support for representational redescription model, suggesting that implicit representations can be redescribed into explicit representations (Karmiloff-Smith, 1986, 1991, 1992, 1994). The redescription-associate learning also suggests that participants need to establish and memorize the associations between explicit representations of perceptual states and their responses. This may be crucial to consolidate explicit representations. Once associated, when similar perceptual states recurred, they would trigger the correct responses associated with them.

Considering that the redescription-associate learning paradigm focused on the dissociation between implicit and explicit processes, it may have clinical implications beyond amusia. As a neurodevelopmental disorder, amusia is akin to other neurodevelopmental disorders such as developmental dyslexia with impairment in reading and congenital prosopagnosia with impairment in face recognition (Couvignou & Kolinsky, 2021; Couvignou et al., 2019, 2023; Mignault Goulet et al., 2012; Peretz, 2016). Specifically, dyslexics or prosopagnosics exhibit intact implicit but impaired explicit processing of phonemes (Bonte & Blomert, 2004; McPherson et al., 1998; Mundy & Carroll, 2012) or faces (Eimer et al., 2012; Rivolta et al., 2012; Stumps et al., 2020) respectively. In other words, these disorders are linked to abnormalities in conscious access to stored implicit knowledge. Based on the similarity of the dissociation between implicit and explicit processes, it seems reasonable to speculate that the redescription-associate learning method might be effective in reducing deficits of awareness in dyslexics and prosopagnosics. However, future research is needed to test these hypotheses.

There are some potential methodological limitations to this study. First, the sample size was relatively small due to the rarity of amusia. It is necessary to conduct additional research with a large sample size in order to verify the efficacy of the redescription-associate learning and the generalizability and the reliability of the findings. Second, although we found the positive effect of music learning on explicit processing of melodic structure for individuals with amusia, this effect reflects near transfer of learning, as similar types of melodic stimuli were used in both the training and testing sessions. Indeed, the scores of the MBEA and pitch threshold might reflect far transfer effect in the present study. However, these data were collected at 10-month follow-up, rather than at posttest and 3-month follow-up after training. Such a manipulation cannot rule out the possibility that far transfer effect may occur earlier, such as at posttest or 3-month follow-up. Therefore, the need for future research to corroborate the near and far transfer effects of the redescription-associate learning on musical structural processing in amusia is warranted.

In conclusion, this study shows that redescription-associate learning may improve explicit processing of musical pitch structure in amusics to a normal level, and this improvement might last up to at least 3 months posttest. These findings provide the first direct evidence of a link between training and neural plasticity in amusics, and suggest that redescription-associate learning may be an effective method to remedy impaired explicit processes at both low-level pitch detection/discrimination and high-level pitch structure processing not only for amusics, but also for those individuals with neurodevelopmental disorders who show preserved implicit but impaired explicit processing.

Declaration of competing interest

None.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31470972, 31500876) and a European Research Council Starting Grant (CAASD, 678733).

References

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., & Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex. *Brain, 136*(5), 1639-1661. https://doi.org/10.1093/brain/awt082

Albouy, P., Peretz, I., Bermudez, P., Zatorre, R. J., Tillmann, B., & Caclin, A. (2019). Specialized neural dynamics for verbal and tonal memory: fMRI evidence in congenital amusia. *Human Brain Mapping, 40*(3), 855-867. https://doi.org/10.1002/hbm.24416

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.).

American Psychological Association. (2020). *Publication manual* of *the American Psychological Association* (7th ed.). https://doi.org/10.1037/0000165-000

Anderson, S., Himonides, E., Wise, K., Welch, G., & Stewart, L. (2012). Is there potential for learning in amusia? A study of the effect of singing intervention in congenital amusia. *Annals of the New York Academy of Sciences, 1252*(1), 345-353. https://doi.org/10.1111/j.1749-6632.2011.06404.x

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125*(2), 238-251. https://doi.org/10.1093/brain/awf028

Barcikowski, R. S., & Robey, R. R. (1984). Decisions in single group repeated measures analysis: Statistical tests and three computer packages. *The American Statistician, 38*(2), 148-150. https://doi.org/10.1080/00031305.1984.10483187

Biderman, N., Luria, R., Teodorescu, A. R., Hajaj, R., & Goshen-Gottstein, Y. (2019). Working memory has better fidelity than long-term memory: The fidelity constraint is not a general property of memory after all. *Psychological Science, 30*(2), 223-237. https://doi.org/10.1177/0956797618813538

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience, 8*, Article 94. https://doi.org/10.3389/fnsys.2014.00094

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics, 58*(1), 125-141. https://doi.org/10.3758/BF03205482

Birmingham, E., Stanley, D., Nair, R., & Adolphs, R. (2015). Implicit social biases in people with autism. *Psychological Science, 26*(11), 1693-1705. https://doi.org/10.1177/0956797615595607

Bischoff-Grethe, A., Hazeltine, E., Bergren, L., Ivry, R. B., & Grafton, S. T. (2009). The influence of feedback valence in associative learning. *NeuroImage, 44*(1), 243-251. https://doi.org/10.1016/j.neuroimage.2008.08.038

Bonte, M. L., & Blomert, L. (2004). Developmental dyslexia: ERP correlates of anomalous phonological processing during spoken word recognition. *Cognitive Brain Research, 21*(3), 360-376. https://doi.org/10.1016/j.cogbrainres.2004.06.010

Brattico, E., Tervaniemi, M., Näätänen, R., & Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Research, 1117*(1), 162-174. https://doi.org/10.1016/j.brainres.2006.08.023

Brown, S. A., & Alibali, M. W. (2018). Promoting strategy change: Mere exposure to alternative strategies helps, but feedback can hurt. *Journal of Cognition and Development, 19*(3), 301-324. https://doi.org/10.1080/15248372.2018.1477778

Brown-Iannuzzi, J. L., Lundberg, K. B., Kay, A. C., & Payne, B. K. (2015). Subjective status shapes political preferences. *Psychological Science, 26*(1), 15-26. https://doi.org/10.1177/0956797614553947

Budd, M. (1995). *Values of art: Pictures, poetry and music*. Penguin Books.

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology, 105*, 290-298. https://doi.org/10.1037/a0031026

Calin-Jageman, R. J., & Cumming, G. (2019a). Estimation for better inference in neuroscience. *eNeuro, 6*(4), Article ENEURO.0205-19.2019. https://doi.org/10.1523/eneuro.0205-19.2019

Calin-Jageman, R. J., & Cumming, G. (2019b). The new statistics for better science: Ask how much, how uncertain, and what else is known. *The American Statistician, 73*(Suppl. 1), 271-280. https://doi.org/10.1080/00031305.2018.1518266

Caplan, J. S., Giménez, D., Hirmas, D. R., Brunsell, N. A., Blair, J. M., & Knapp, A. K. (2019). Decadal-scale shifts in soil hydraulic properties as induced by altered precipitation. *Science Advances, 5*(9), Article eaau6635. https://doi.org/10.1126/sciadv.aau6635

Carrión, R. E., & Bly, B. M. (2008). The effects of learning on event-related potential correlates of musical expectancy. *Psychophysiology, 45*(5), 759-775. https://doi.org/10.1111/j.1469-8986.2008.00687.x

Cermeño-Aínsa, S. (2021). The perception/cognition distincton: Challenging the representational account. *Consciousness and Cognition, 95*, Article 103216. https://doi.org/10.1016/j.concog.2021.103216

Cheung, S. F., Pesigan, I. J. A., & Vong, W. N. (2022). 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*. Advance online publication. https://doi.org/10.3758/s13428-022-01808-5

Cleeremans, A. (2019). Consciousness (unconsciously) designs itself. *Journal of Consciousness Studies, 26*(3-4), 88-111.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*(12), 1304–1312. https://doi.org/10.1037/0003-066X.45.12.1304

Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychological Review, 121*(1), 33-65. https://doi.org/10.1037/a0034695

Couvignou, M., & Kolinsky, R. (2021). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia in children. *Neuropsychologia, 155*, Article 107811. https://doi.org/10.1016/j.neuropsychologia.2021.107811

Couvignou, M., Peretz, I., & Ramus, F. (2019). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia. *Cognitive Neuropsychology, 36*(1-2), 1-17. https://doi.org/10.1080/02643294.2019.1578205

Couvignou, M., Tillmann, B., Caclin, A., & Kolinsky, R. (2023). Do developmental dyslexia and congenital amusia share underlying impairments? *Child Neuropsychology*. Advance online publication. https://doi.org/10.1080/09297049.2022.2162031

Cumming, G. (2013). The new statistics: A how- to guide. *Australian Psychologist, 48*(3), 161-170. https://doi.org/10.1111/ap.12018

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1), 7-29. https://doi.org/10.1177/0956797613504966

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3*(2), 255-267. https://doi.org/10.1016/0010-0285(72)90007-2

Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA. *International Review of Social Psychology, 32*(1), Article 13. https://doi.org/10.5334/irsp.198

Eimer, M., Gosling, A., & Duchaine, B. (2012). Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain, 135*(2), 542-554. https://doi.org/10.1093/brain/awr347

Elkin, L. A., Kay, M., Higgins, J. J., & Wobbrock, J. O. (2021). An aligned rank transform procedure for multifactor contrast tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (pp. 754- 768). Association for Computing Machinery. https://doi.org/10.1145/3472749.3474784

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist, 63*(7), 591-601. https://doi.org/10.1037/0003-066X.63.7.591

Feys, J. (2016). New nonparametric rank tests for interactions in factorial designs with repeated measures. *Journal of Modern Applied Statistical Methods, 15*(1), 78-99. https://doi.org/10.22237/jmasm/1462075500

Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Fiveash, A., Thompson, W. F., Badcock, N. A., & McArthur, G. (2018). Syntactic processing in music and language: Effects of interrupting auditory streams with alternating timbres. *International Journal of Psychophysiology, 129*, 31-40. https://doi.org/10.1016/j.ijpsycho.2018.05.003

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain, 127*(4), 801-810. https://doi.org/10.1093/brain/awh105

Graves, J. E., Pralus, A., Fornoni, L., Oxenham, A. J., Caclin, A., & Tillmann, B. (2019). Short- and long-term memory for pitch and non-pitch contours: Insights from congenital amusia. *Brain and Cognition, 136*, Article 103614. https://doi.org/10.1016/j.bandc.2019.103614

Griffiths, P., & Needleman, J. (2019). Statistical significance testing and p-values: Defending the indefensible? A discussion paper and position statement. *International Journal of Nursing Studies, 99*, Article 103384. https://doi.org/10.1016/j.ijnurstu.2019.07.001

Gruhn, W. (2005). Understanding musical understanding. In D. J. Elliot (Ed.), *Praxial music education: Reflections and dialogues* (pp. 98-111). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195385076.003.05

Henry, M. J., & McAuley, J. D. (2010). On the prevalence of congenital amusia. *Music Perception, 27*(5), 413-418. https://doi.org/10.1525/mp.2010.27.5.413

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15*(5), 356-360. https://doi.org/10.1111/j.0956-7976.2004.00683.x

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2011). Fine-grained pitch discrimination in congenital amusics with Mandarin Chinese. *Music Perception, 28*(5), 519-526. https://doi.org/10.1525/mp.2011.28.5.519

Jiang, C., Lim, V. K., Wang, H., & Hamm, J. P. (2013). Difficulties with pitch discrimination influences pitch memory performance: Evidence from congenital amusia. *PLOS ONE, 8*(10), Article e79216. https://doi.org/10.1371/journal.pone.0079216

Jiang, C., Liu, F., & Thompson, W. F. (2016). Impaired explicit processing of musical syntax and tonality in a group of Mandarin-speaking congenital amusics. *Music Perception, 33*(4), 401-413. https://doi.org/10.1525/mp.2016.33.4.401

Jiang, C., Liu, F., & Wong, P. C. M. (2017). Sensitivity to musical emotion is influenced by tonal structure in congenital amusia. *Scientific Reports, 7*(1), Article 7624. https://doi.org/10.1038/s41598-017-08005-x

Kalmus, H., & Fry, D. B. (1980). On tune deafness (dysmelodia): Frequency, development, genetics and musical background. *Annals of Human Genetics, 43*(4), 369-382. https://doi.org/10.1111/j.1469-1809.1980.tb01571.x

Kappenman, E. S., & Luck, S. J. (2016). Best practices for event-related potential research in clinical populations. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1*(2), 110-115. https://doi.org/10.1016/j.bpsc.2015.11.007

Karadaghy, O. A., Hong, H., Scott-Wittenborn, N., Sinha, P., Suko, J., Tait, S., Wamkpah, N. S., Kallogjeri, D., & Piccirillo, J. F. (2017). Reporting of effect size and confidence intervals in JAMA Otolaryngology–Head & Neck Surgery. *JAMA Otolaryngology–Head & Neck Surgery, 143*(11), 1075-1080. https://doi.org/10.1001/jamaoto.2017.1504

Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition, 23*(2), 95-147. https://doi.org/10.1016/0010-0277(86)90040-5

Karmiloff-Smith, A. (1991). Beyond modularity: Innate constraints and developmental change. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 171-197). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315807805-15

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental prespective on cognitive science*. MIT Press. https://doi.org/10.7551/mitpress/1579.001.0001

Karmiloff-Smith, A. (1994). Précis of Beyond modularity: A developmental perspective on cognitive science. *Behavioral and Brain Sciences, 17*(4), 693-707. https://doi.org/10.1017/S0140525X00036621

Kertz-Welzel, A. (2005). The Pied Piper of Hamelin: Adorno on music education. *Research Studies in Music Education, 25*(1), 1-12. https://doi.org/10.1177/1321103X050250010301

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*(5), 746-759. https://doi.org/10.1177/0013164496056005002

Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). American Psychological Association. https://doi.org/10.1037/14136-000

Koelsch, S. (2012). *Brain and music*. John Wiley & Sons.

Koelsch, S., Gunter, T. C., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of Cognitive Neuroscience, 12*(3), 520-541. https://doi.org/10.1162/089892900562183

Koelsch, S., & Jentschke, S. (2008). Short-term effects of processing musical syntax: An ERP study. *Brain Research, 1212*, 55-62. https://doi.org/10.1016/j.brainres.2007.10.078

Koelsch, S., Jentschke, S., Sammler, D., & Mietchen, D. (2007). Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology, 44*(3), 476-490. https://doi.org/10.1111/j.1469-8986.2007.00517.x

Koelsch, S., Maess, B., Grossmann, T., & Friederici, A. D. (2003). Electric brain responses reveal gender differences in music processing. *NeuroReport, 14*(5), 709-713. https://doi.org/10.1097/01.wnr.0000065762.60383.67

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences, 110*(38), 15443-15448. https://doi.org/10.1073/pnas.1300272110

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195148367.001.0001

Krumhansl, C. L., & Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music. *Memory & Cognition, 10*(3), 243-251. https://doi.org/10.3758/BF03197636

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*(4), 334-368. https://doi.org/10.1037/0033-295X.89.4.334

Kubit, B. M., & Janata, P. (2022). Spontaneous mental replay of music improves memory for incidentally associated event knowledge. *Journal of Experimental Psychology: General, 151*(1), 1-24. https://doi.org/10.1037/xge0001050

Kuklick, L., & Lindner, M. A. (2021). Computer-based knowledge of results feedback in different delivery modes: Effects on performance, motivation, and achievement emotions. *Contemporary Educational Psychology, 67*, Article 102001. https://doi.org/10.1016/j.cedpsych.2021.102001

Lantz, M. E., Cui, A.-X., & Cuddy, L. L. (2020). The role of duration and frequency of occurrence in perceived pitch structure. *PLOS ONE, 15*(9), Article e0239582. https://doi.org/10.1371/journal.pone.0239582

Lehne, M., Rohrmeier, M., Gollmann, D., & Koelsch, S. (2013). The influence of different structural features on felt musical tension in two piano pieces by Mozart and Mendelssohn. *Music Perception, 31*(2), 171-185. https://doi.org/10.1525/mp.2013.31.2.171

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception, 17*(4), 481-509. https://doi.org/10.2307/40285830

Lévêque, Y., Lalitte, P., Fornoni, L., Pralus, A., Albouy, P., Bouchet, P., Caclin, A., & Tillmann, B. (2022). Tonal structures benefit short-term memory for real music:

Evidence from non-musicians and individuals with congenital amusia. *Brain and Cognition, 161*, Article 105881. https://doi.org/10.1016/j.bandc.2022.105881

Lima, J. P. d., Iervolino, S. M. S., & Schochat, E. (2018). Musical and temporal auditory skills in cochlear implant users after music therapy. *Codas, 30*(6), Article e20180006. https://doi.org/10.1590/2317-1782/20182018006

Liu, F., Jiang, C., Francart, T., Chan, A. H. D., & Wong, P. C. M. (2017). Perceptual learning of pitch direction in congenital amusia. *Music Perception, 34*(3), 335. https://doi.org/10.1525/mp.2017.34.3.335

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain, 133*(6), 1682-1693. https://doi.org/10.1093/brain/awq089

Liu, F., Xu, Y., Patel, A. D., Francart, T., & Jiang, C. (2012). Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers. *Brain and Cognition, 79*(3), 209-215. https://doi.org/10.1016/j.bandc.2012.03.008

Liu, H. (2015). *Comparing Welch ANOVA, a Kruskal-Wallis test, and traditional ANOVA in case of heterogeneity of variance* (Publication No. 1601507) [Master's thesis, Virginia Commonwealth University]. ProQuest Dissertations and Theses Global.

Loui, P. (2016). Disorders of music cognition. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 307-323). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.22

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *The Journal of Neuroscience, 29*(33), 10215-10220. https://doi.org/10.1523/JNEUROSCI.1701-09.2009

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology, 18*(8), R331-R332. https://doi.org/https://doi.org/10.1016/j.cub.2008.02.045

Lu, X., Ho, H. T., Sun, Y., Johnson, B. W., & Thompson, W. F. (2016). The influence of visual information on auditory processing in individuals with congenital amusia: An ERP study. *NeuroImage, 135*, 142-151. https://doi.org/10.1016/j.neuroimage.2016.04.043

Lu, X., Sun, Y., Ho, H. T., & Thompson, W. F. (2017). Pitch contour impairment in congenital amusia: New insights from the Self-paced Audio-visual Contour Task (SACT). *PLOS ONE, 12*(6), Article e0179252. https://doi.org/10.1371/journal.pone.0179252

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146-157. https://doi.org/10.1111/psyp.12639

Maimon, N. B., Lamy, D., & Eitan, Z. (2022). Do Picardy thirds smile? Tonal hierarchy and tonal valence: Explicit and implicit measures. *Music Perception, 39*(5), 443-467. https://doi.org/10.1525/mp.2022.39.5.443

Mandell, J., Schulze, K., & Schlaug, G. (2007). Congenital amusia: An auditory-motor feedback disorder? *Restorative Neurology and Neuroscience, 25*(3-4), 323-334.

Marmel, F., Perrin, F., & Tillmann, B. (2011). Tonal expectations influence early pitch processing. *Journal of Cognitive Neuroscience, 23*(10), 3095-3104. https://doi.org/10.1162/jocn.2011.21632

Marmel, F., Tillmann, B., & Delbé, C. (2010). Priming in melody perception: Tracking down the strength of cognitive expectations. *Journal of Experimental Psychology: Human Perception and Performance, 36*(4), 1016-1028. https://doi.org/10.1037/a0018735

Marsh, E. J., Lozito, J. P., Umanath, S., Bjork, E. L., & Bjork, R. A. (2012). Using verification feedback to correct errors made on a multiple-choice test. *Memory, 20*(6), 645-653. https://doi.org/10.1080/09658211.2012.684882

McPherson, W. B., Ackerman, P. T., Holcomb, P. J., & Dykman, R. A. (1998). Event-related brain potentials elicited during phonological processing differentiate subgroups of reading disabled adolescents. *Brain and Language, 62*(2), 163-185. https://doi.org/10.1006/brln.1997.1893

Mignault Goulet, G., Moreau, P., Robitaille, N., & Peretz, I. (2012). Congenital amusia persists in the developing brain after daily music listening. *PLOS ONE, 7*(5), Article e36860. https://doi.org/10.1371/journal.pone.0036860

Moreau, P., Jolicœur, P., & Peretz, I. (2009). Automatic brain responses to pitch changes in congenital amusia. *Annals of the New York Academy of Sciences, 1169*(1), 191-194. https://doi.org/10.1111/j.1749-6632.2009.04775.x

Moreau, P., Jolicœur, P., & Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials. *Brain and Cognition, 81*(3), 337-344. https://doi.org/10.1016/j.bandc.2013.01.004

Mundy, I. R., & Carroll, J. M. (2012). Speech prosody and developmental dyslexia: Reduced phonological awareness in the context of intact phonological representations. *Journal of Cognitive Psychology, 24*(5), 560-581. https://doi.org/10.1080/20445911.2012.662341

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews, 82*(4), 591-605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain, 133*(9), 2635-2642. https://doi.org/10.1093/brain/awq178

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging, 22*(1), 202-208. https://doi.org/10.1037/0882-7974.22.1.202

Nelson, K. (1996). *Language in cognitive development: The emergence of the mediated mind*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174619

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113. https://doi.org/10.1016/0028-3932(71)90067-4

Olson, M. H., & Ramirez, J. J. (2020). *An introduction to theories of learning* (10th ed.). Routledge. https://doi.org/10.4324/9781003014447

Omigie, D., Pearce, M. T., & Stewart, L. (2012). Tracking of pitch probabilities in congenital amusia. *Neuropsychologia, 50*(7), 1483-1493. https://doi.org/10.1016/j.neuropsychologia.2012.02.034

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Park, H. S. (2013). The development and the effects of verbalization on representational redescription in children's drawings. *Korean Journal of Child Studies, 34*(6), 139-158. https://doi.org/10.5723/kjcs.2013.34.6.139

Park, H. S. (2015). Relationships between verbalization through decomposing- connecting object representation and representational redescription process in children's drawing behavior. *Journal of Children's Media & Education, 14*(3), 83-105.

Park, H. S., & Choi, K.-S. (2006). The effects of verbalization on representational redescription in a balance beam task. *The Korean Journal of Developmental Psychology, 19*(1), 69-87.

Park, H. S., Jang, S. H., & Choi, K. S. (2008). The effects of verbalization on the representational redescription of number conservation. *Korean Journal of Early Childhood Education, 28*(6), 119-141. https://doi.org/10.18023/kjece.2008.28.6.006

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674-681. https://doi.org/10.1038/nn1082

Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195123753.001.0001

Pavlov, I. P. (1955). *I. P. Pavlov: Selected works* (K. S. Koshtoyants, Ed.). Foreign Languages Publishing House.

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences, 112*(33), 10089-10092. https://doi.org/10.1073/pnas.1504933112

Perdices, M. (2018). Null hypothesis significance testing, p-values, effects sizes and confidence intervals. *Brain Impairment, 19*(1), 70-80. https://doi.org/10.1017/BrImp.2017.28

Peretz, I. (2016). Neurobiology of congenital amusia. *Trends in Cognitive Sciences, 20*(11), 857-867. https://doi.org/10.1016/j.tics.2016.09.002

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. t. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33*(2), 185-191. https://doi.org/10.1016/S0896-6273(01)00580-3

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: In tune, out of key, and unaware. *Brain, 132*(5), 1277-1286. https://doi.org/10.1093/brain/awp055

Peretz, I., Brattico, E., & Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Annals of Neurology, 58*(3), 478-482. https://doi.org/10.1002/ana.20606

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences, 999*(1), 58-75. https://doi.org/10.1196/annals.1284.006

Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., & Béland, R. (2013). A novel tool for evaluating children's musical abilities across age and culture. *Frontiers in Systems Neuroscience, 7*, Article 30. https://doi.org/10.3389/fnsys.2013.00030

Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics, 25*, 625–630. https://doi.org/10.1038/ejhg.2017.15

Pine, K. J., & Messer, D. J. (1998). Group collaboration effects and the explicitness of children's knowledge. *Cognitive Development, 13*(1), 109-126. https://doi.org/10.1016/S0885-2014(98)90023-5

Quilty-Dunn, J. (2020). Perceptual pluralism. *Noûs, 54*(4), 807-838. https://doi.org/10.1111/nous.12285

Quiroga-Martinez, D. R., Tillmann, B., Brattico, E., Cholvy, F., Fornoni, L., Vuust, P., & Caclin, A. (2021). Listeners with congenital amusia are sensitive to context uncertainty in melodic sequences. *Neuropsychologia, 158*, Article 107911. https://doi.org/10.1016/j.neuropsychologia.2021.107911

Rivolta, D., Palermo, R., Schmalzl, L., & Coltheart, M. (2012). Covert face recognition in congenital prosopagnosia: A group study. *Cortex, 48*(3), 344-352. https://doi.org/10.1016/j.cortex.2011.01.005

Samson, D. R., Vining, A., & Nunn, C. L. (2019). Sleep influences cognitive performance in lemurs. *Animal Cognition, 22*(5), 697-706. https://doi.org/10.1007/s10071-019-01266-1

Sauder, D. C., & DeMars, C. E. (2019). An updated recommendation for multiple comparisons. *Advances in Methods and Practices in Psychological Science, 2*(1), 26-44. https://doi.org/10.1177/2515245918808784

Saunders, B., Riesel, A., Klawohn, J., & Inzlicht, M. (2018). Interpersonal touch enhances cognitive control: A neurophysiological investigation. *Journal of Experimental Psychology: General, 147*(7), 1066-1077. https://doi.org/10.1037/xge0000412

Schober, P., Bossers, S. M., & Schwarte, L. A. (2018). Statistical significance versus clinical importance of observed effect sizes: What do P values and confidence intervals really represent? *Anesthesia & Analgesia, 126*(3), 1068-1072. https://doi.org/10.1213/ane.0000000000002798

Sears, D. R., Pearce, M. T., Spitzer, J., Caplin, W. E., & McAdams, S. (2019). Expectations for tonal cadences: Sensory and cognitive priming effects. *Quarterly Journal of Experimental Psychology, 72*(6), 1422-1438. https://doi.org/10.1177/1747021818814472

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology, 23*(6), 695-703. https://doi.org/10.1111/j.1469-8986.1986.tb00696.x

Skinner, B. F. (1953). *Science and human behavior*. Macmillan.

Skinner, B. F. (1958). Teaching machines. *Science, 128*(3330), 969-977. https://doi.org/10.1126/science.128.3330.969

Sohn, D. (1982). Sex differences in achievement self-attributions: An effect-size analysis. *Sex Roles, 8*(4), 345-357. https://doi.org/10.1007/BF00287274

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*(2), 164-182. https://doi.org/10.1037/1082-989X.9.2.164

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience, 18*(8), 1380-1393. https://doi.org/10.1162/jocn.2006.18.8.1380

Streiner, D. L. (2015). Best (but oft-forgotten) practices: the multiple problems of multiplicity—whether and how to correct for many statistical tests. *The American Journal of Clinical Nutrition, 102*(4), 721-728. https://doi.org/10.3945/ajcn.115.113548

Stumps, A., Saad, E., Rothlein, D., Verfaellie, M., & DeGutis, J. (2020). Characterizing developmental prosopagnosia beyond face perception: Impaired recollection but intact familiarity recognition. *Cortex, 130*, 64-77. https://doi.org/10.1016/j.cortex.2020.04.016

Sun, L., Thompson, W. F., Liu, F., Zhou, L., & Jiang, C. (2020). The human brain processes hierarchical structures of meter and harmony differently: Evidence from musicians and nonmusicians. *Psychophysiology, 57*(9), Article e13598. https://doi.org/10.1111/psyp.13598

Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital amusias. In M. J. Aminoff, F. Boller, & D. F. Swaab (Eds.), *Handbook of Clinical Neurology* (Vol. 129, pp. 589-605). Elsevier. https://doi.org/10.1016/B978-0-444-62630-1.00033-0

Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia: Evidence from a gating paradigm. *Cortex, 59*, 84-94. https://doi.org/10.1016/j.cortex.2014.07.012

Tillmann, B., Gosselin, N., Bigand, E., & Peretz, I. (2012). Priming paradigm reveals harmonic structure processing in congenital amusia. *Cortex, 48*(8), 1073-1078. https://doi.org/10.1016/j.cortex.2012.01.001

Tillmann, B., Lalitte, P., Albouy, P., Caclin, A., & Bigand, E. (2016). Discrimination of tonal and atonal music in congenital amusia: The advantage of implicit tasks. *Neuropsychologia, 85*, 10-18. https://doi.org/10.1016/j.neuropsychologia.2016.02.027

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research, 1640*(Part B), 251-263. https://doi.org/10.1016/j.brainres.2015.10.035

Tillmann, B., Poulin-Charronnat, B., Gaudrain, E., Akhoun, I., Delbé, C., Truy, E., & Collet, L. (2019). Implicit processing of pitch in postlingually deafened cochlear implant users. *Frontiers in Psychology, 10*, Article 1990. https://doi.org/10.3389/fpsyg.2019.01990

Vuvan, D. T., Paquette, S., Mignault Goulet, G., Royal, I., Felezeu, M., & Peretz, I. (2018). The Montreal protocol for identification of amusia. *Behavior Research Methods, 50*(2), 662-672. https://doi.org/10.3758/s13428-017-0892-8

Wang, J., Zhang, C., Wan, S., & Peng, G. (2017). Is congenital amusia a disconnection syndrome? A study combining tract- and network-based analysis. *Frontiers in Human Neuroscience, 11*, Article 473. https://doi.org/10.3389/fnhum.2017.00473

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician, 70*(2), 129-133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "$p < 0.05$". *The American Statistician, 73*(Suppl. 1), 1-19. https://doi.org/10.1080/00031305.2019.1583913

Weiss, M. W., & Peretz, I. (2022). Improvisation is a novel tool to study musicality. *Scientific Reports, 12*(1), Article 12595. https://doi.org/10.1038/s41598-022-15312-5

Whiteford, K. L., & Oxenham, A. J. (2017). Auditory deficits in amusia extend beyond poor pitch perception. *Neuropsychologia, 99*, 213-224. https://doi.org/10.1016/j.neuropsychologia.2017.03.018

Whiteford, K. L., & Oxenham, A. J. (2018). Learning for pitch and melody discrimination in congenital amusia. *Cortex, 103*, 164-178. https://doi.org/10.1016/j.cortex.2018.03.012

Wilbiks, J. M. P., Vuvan, D. T., Girard, P.-Y., Peretz, I., & Russo, F. A. (2016). Effects of vocal training in a musicophile with congenital amusia. *Neurocase, 22*(6), 526-537. https://doi.org/10.1080/13554794.2016.1263339

Wilcox, R. R. (2022). *Introduction to robust estimation and hypothesis testing* (5th ed.). Elsevier. https://doi.org/10.1016/C2019-0-01225-3

Wilcox, R. R., & Serang, S. (2017). Hypothesis testing, p values, confidence intervals, measures of effect size, and Bayesian methods in light of modern robust techniques. *Educational and Psychological Measurement, 77*(4), 673-689. https://doi.org/10.1177/0013164416667983

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 143–146). Association for Computing Machinery. https://doi.org/10.1145/1978942.1978963

Wobbrock, J. O., & Kay, M. (2016). Nonparametric statistics in human–computer interaction. In J. Robertson & M. Kaptein (Eds.), *Modern statistical methods for HCI* (pp. 135-170). Springer. https://doi.org/10.1007/978-3-319-26633-6_7

Zendel, B. R., Lagrois, M.-É., Robitaille, N., & Peretz, I. (2015). Attending to pitch information inhibits processing of pitch information: The curious case of amusia. *The Journal of Neuroscience, 35*(9), 3815-3824. https://doi.org/10.1523/JNEUROSCI.3766-14.2015

Zhou, L., Liu, F., Jiang, J., Jiang, H., & Jiang, C. (2019). Abnormal neural responses to harmonic syntactic structures in congenital amusia. *Psychophysiology, 56*(9), Article e13394. https://doi.org/10.1111/psyp.13394

Table 1

Demographic and Diagnostic Information for the Three Groups

Measure	Trained amusics (n = 8)	Untrained amusics (n = 8)	Controls (n = 11)	Group difference
Demographic variable				
Age (years)	22.75 ± 1.58	23.63 ± 2.26	23.27±1.56	$F(2, 24) = 0.48$, $p = .624$, $\omega_p^2 = -.040$, 95% CI [.000, .000]
Sex (male/female)	2/6	3/5	3/8	p = 1.000, Fisher's exact test
Education (years)	15.63 ± 1.41	16.38 ± 2.07	16.64 ± 1.43	$F(2, 24) = 0.91$, $p = .415$, $\omega_p^2 = -.007$, 95% CI [.000, .000]
MBEA				
Scale	19.25 ± 2.38	16.63 ± 2.72	27.27 ± 1.56	$F(2, 24) = 61.85$, $p < .001$, $\omega_p^2 = .818$, 95% CI [.657, .889]
Contour	19.00 ± 3.16	19.00 ± 3.46	27.18 ± 1.08	$F(2, 10.65) = 39.99$, $p < .001$, $\omega_p^2 = .851$, 95% CI [.579, .926]
Interval	18.00 ± 2.98	17.38 ± 3.96	27.45 ± 1.57	$F(2, 11.74) = 47.60$, $p < .001$, $\omega_p^2 = .863$, 95% CI [.634, .930]
Rhythm	20.75 ± 4.33	19.50 ± 4.66	27.36 ± 2.01	$F(2, 11.60) = 15.18$, $p < .001$, $\omega_p^2 = .660$, 95% CI [.209, .823]
Meter	16.38 ± 3.02	19.38 ± 5.60	27.09 ± 1.97	$F(2, 12.13) = 39.45$, $p < .001$, $\omega_p^2 = .836$, 95% CI [.574, .915]
Memory	20.75 ± 3.58	20.38 ± 5.34	27.82 ± 1.89	$F(2, 11.58) = 17.10$, $p < .001$, $\omega_p^2 = .688$, 95% CI [.255, .839]
Melodic score	56.25 ± 5.65	53.00 ± 7.46	81.91 ± 3.18	$F(2, 12.00) = 99.26$, $p < .001$, $\omega_p^2 = .929$, 95% CI [.808, .964]
Global score (%)	63.40 ± 6.15	62.36 ± 6.52	91.21 ± 2.43	$F(2, 11.09) = 124.41$, $p < .001$, $\omega_p^2 = .946$, 95% CI [.846, .973]
Pitch perception threshold (semitones)				

Pitch change detection	1.81 ± 1.65	1.10 ± 1.12	0.45 ± 0.30	$F(2, 10.02) = 3.56$, $p = .068$, $\omega_p^2 = .282$, 95% CI [.000, .605]
Pitch direction discrimination	5.48 ± 1.83	4.54 ± 1.53	2.12 ± 2.15	$F(2, 15.65) = 6.96$, $p = .007$, $\omega_p^2 = .390$, 95% CI [.008, .638]

Note. The maximum score is 30 for each MBEA subtest, and 90 for the melodic score that is the sum of the scores on the scale, contour, and interval subtests. Values are $M \pm SD$. Because the data on age, education, and scale were normally distributed and had equal variances, the classic Fisher's *F*-test was used; but the data on other measures violated any assumption, the Welch's *F*-test was used (Delacre et al., 2019; Liu, 2015).

Table 2

ANOVA Results for the Effects of Time, Group, Regularity and Hemisphere on the ERAN (150–250 ms)

Effect	*F*	*df*	*p*	ω_p^2	95% CI
Time	7.24	1, 24	.013	.194	[.000, .452]
Group	0.64	2, 24	.536	−.027	[.000, .000]
Regularity	13.43	1, 24	.001	.323	**[.055, .561]**
Hemisphere	4.42	1, 24	.046	.116	[.000, .373]
Time × Group	1.76	2, 24	.193	.053	[.000, .254]
Time × Regularity	3.63	1, 24	.069	.092	[.000, .344]
Time × Hemisphere	0.20	1, 24	.658	−.032	[.000, .000]
Group × Regularity	5.34	2, 24	.012	.243	[.000, .482]
Group × Hemisphere	1.89	2, 24	.173	.062	[.000, .269]
Regularity × Hemisphere	2.46	1, 24	.130	.053	[.000, .291]
Time × Group × Regularity	7.45	2, 24	.003	.323	**[.030, .551]**
Time × Group × Hemisphere	1.38	2, 24	.271	.027	[.000, .193]
Time × Regularity × Hemisphere	0.23	1, 24	.633	−.031	[.000, .000]
Group × Regularity × Hemisphere	3.90	2, 24	.034	.177	[.000, .418]
Time × Group × Regularity × Hemisphere	1.36	2, 24	.276	.026	[.000, .190]

Table 3

ANOVA Results for the Effects of Time, Regularity and Hemisphere on the ERAN (150–250 ms) in Trained Amusics

Effect	$F(1, 6)$	p	ω_p^2	95% CI
Time	3.43×10^{-6}	.999	−.143	[.000, .000]
Regularity	22.98	.003	.733	**[.146, .890]**
Hemisphere	1.98	.209	.109	[.000, .573]
Time ×Regularity	2.12	.196	.123	[.000, .584]
Time × Hemisphere	0.08	.788	−.130	[.000, .000]
Regularity × Hemisphere	0.03	.870	−.138	[.000, .000]
Time × Regularity × Hemisphere	1.96	.212	.107	[.000, .571]

Table 4

Means, Standard Deviations, and Two-Way ANOVA for MBEA and Pitch Threshold

Measure	Trained amusics		Untrained amusics		ANOVA				
	Pretest	Follow-up	Pretest	Follow-up	Effect	*F*(1, 11)	*p*	ω_p^2	95% CI
MBEA									
Scale	18.71 ± 1.98	20.86 ± 3.53	16.67 ± 2.80	19.67 ± 4.59	Time	5.08	.045	.239	[.000, .584]
					Group	1.24	.290	.018	[.000, .337]
					Time × Group	0.14	.714	−.071	[.000, .000]
Contour	19.29 ± 3.30	19.00 ± 2.24	19.67 ± 3.78	21.33 ± 2.50	Time	0.47	.505	−.043	[.000, .000]
					Group	1.03	.331	.002	[.000, .227]
					Time × Group	0.95	.351	−.004	[.000, .000]
Interval	18.14 ± 3.18	20.57 ± 2.57	17.17 ± 3.92	19.83 ± 3.71	Time	2.89	.117	.127	[.000, .492]
					Group	0.60	.454	−.032	[.000, .000]
					Time × Group	0.01	.938	−.082	[.000, .000]
Rhythm	20.14 ± 4.30	23.43 ± 2.51	19.83 ± 5.23	23.83 ± 1.17	Time	9.27	.011	.389	**[.004, .681]**
					Group	0.001	.977	−.083	[.000, .000]
					Time × Group	0.09	.771	−.075	[.000, .000]
Meter	16.43 ± 3.26	18.43 ± 5.32	19.83 ± 6.05	20.00 ± 3.95	Time	0.56	.471	−.035	[.000, .000]
					Group	1.28	.283	.021	[.000, .346]
					Time × Group	0.40	.540	−.048	[.000, .000]
Memory	20.14 ± 3.39	25.57 ± 2.57	20.00 ± 6.26	25.83 ± 3.43	Time	29.16	< .001	.684	**[.270, .841]**
					Group	0.001	.977	−.083	[.000, .000]
					Time × Group	0.04	.850	−.080	[.000, .000]
Global score (%)	62.70 ± 6.28	71.03 ± 4.41	62.87 ± 5.68	72.50 ± 7.25	Time	21.24	< .001	.609	**[.167, .802]**
					Group	0.09	.765	−.075	[.000, .000]
					Time × Group	0.11	.746	−.073	[.000, .000]
Pitch perception threshold (semitones)									

Pitch change detection	1.67 ± 1.73	0.91 ± 0.52	1.18 ± 1.30	0.68 ± 0.76	Time	1.74	.214	.054	[.000, .409]
					Group	1.27	.284	.020	[.000, .344]
					Time × Group	0.01	.924	−.082	[.000, .000]
Pitch direction discrimination	5.19 ± 1.76	2.99 ± 1.62	4.60 ± 1.23	4.24 ± 2.64	Time	7.77	.018	.342	[.000, .653]
					Group	0.12	.732	−.073	[.000, .000]
					Time × Group	4.06	.069	.191	[.000, .547]

Note. Values are $M \pm SD$. The data on pitch change detection violated normality, the ANOVA thus was based on the aligned-and-ranked data. However, the ANOVAs on other measures were based on the original data due to normality and homogeneity of variances.

Figure 1

Examples of Melodies With a Regular and an Irregular Ending

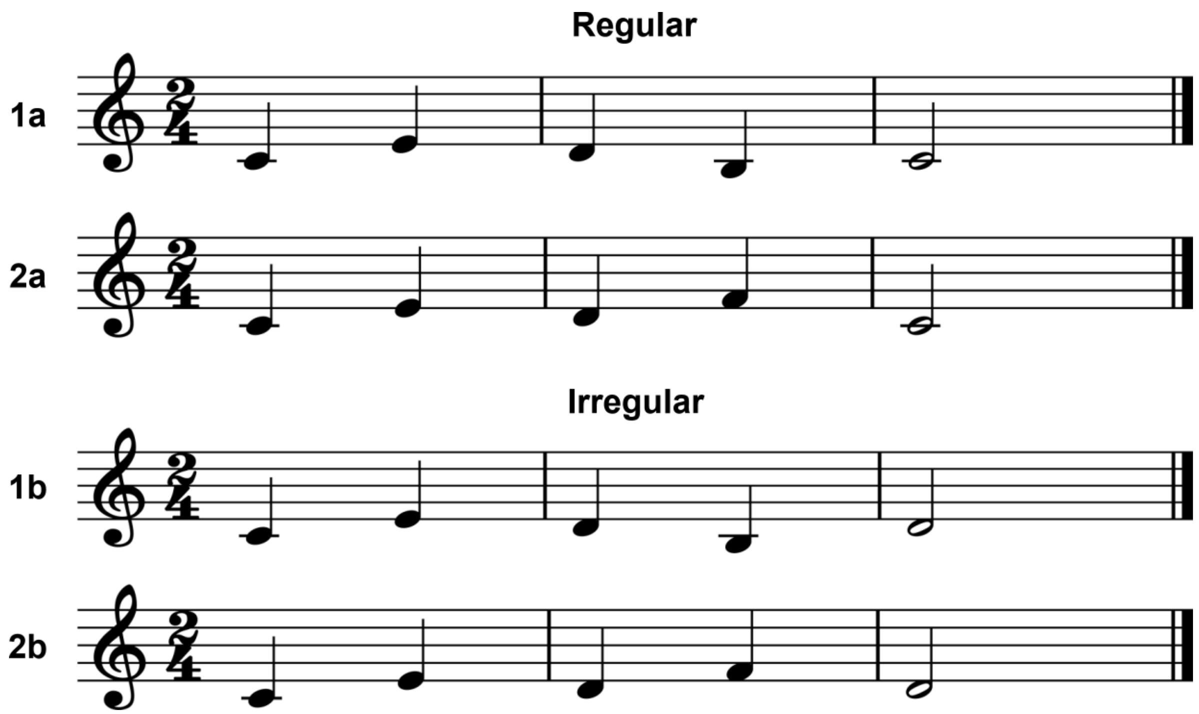


Figure 2

Timeline of Testing and Training

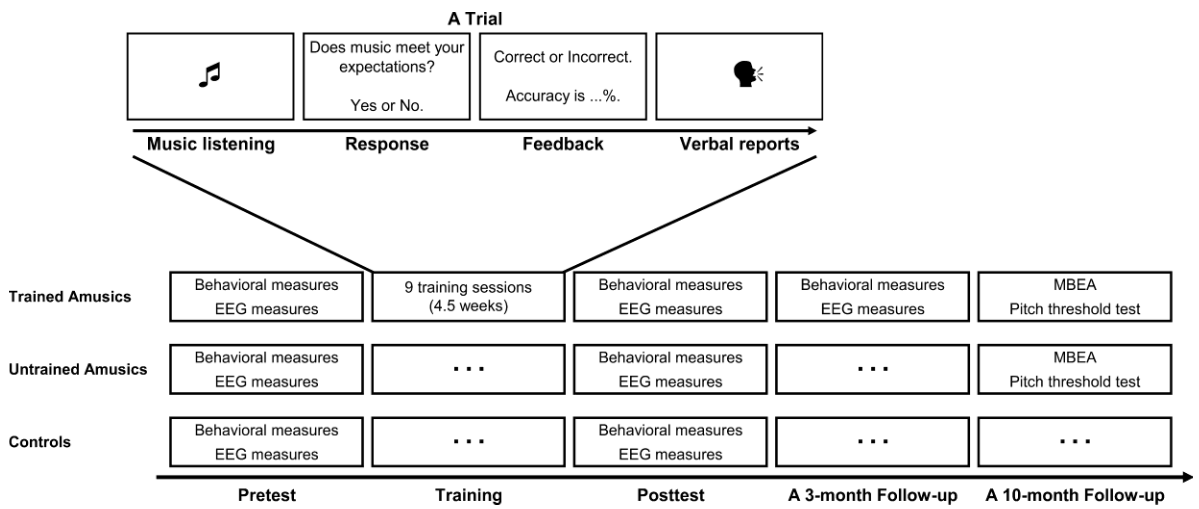
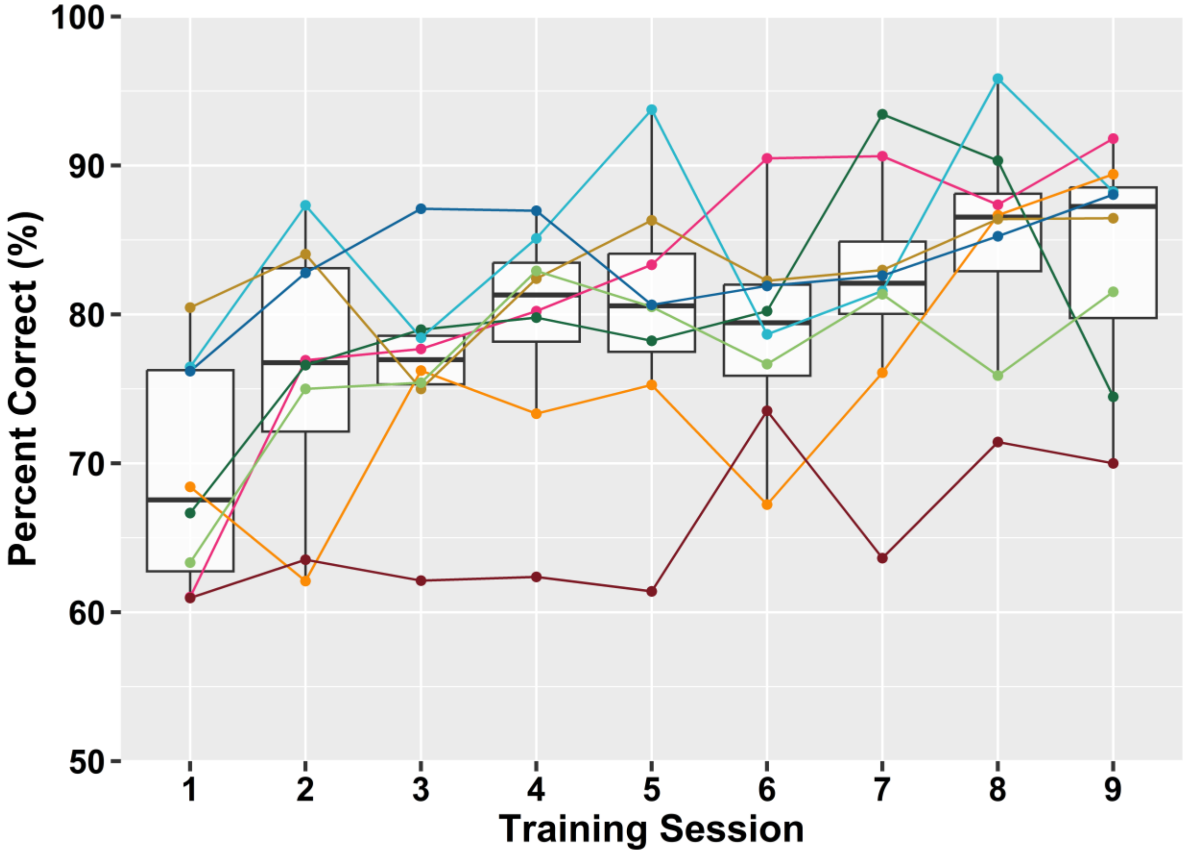


Figure 3

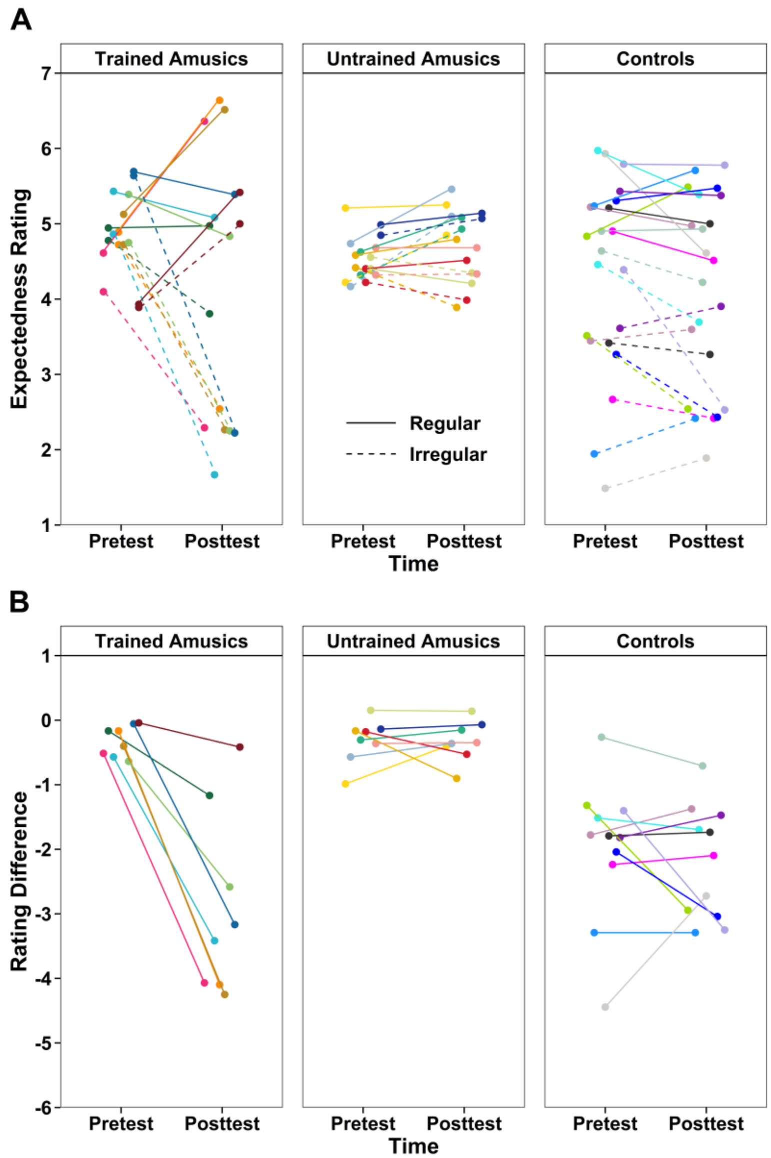
Percentage of Correct Responses Across the Nine Training Sessions for Trained Amusics



Note. Each data point represents an individual sample. These data are summarized in boxplots in which the bisecting line marks the median, the box signifies the upper and lower quartiles, and the whiskers represent the minimum and maximum within 1.5 times the interquartile range. The graph was plotted with the GGally (Version 2.1.2) and ggplot2 (Version 3.4.0) packages within the R.

Figure 4

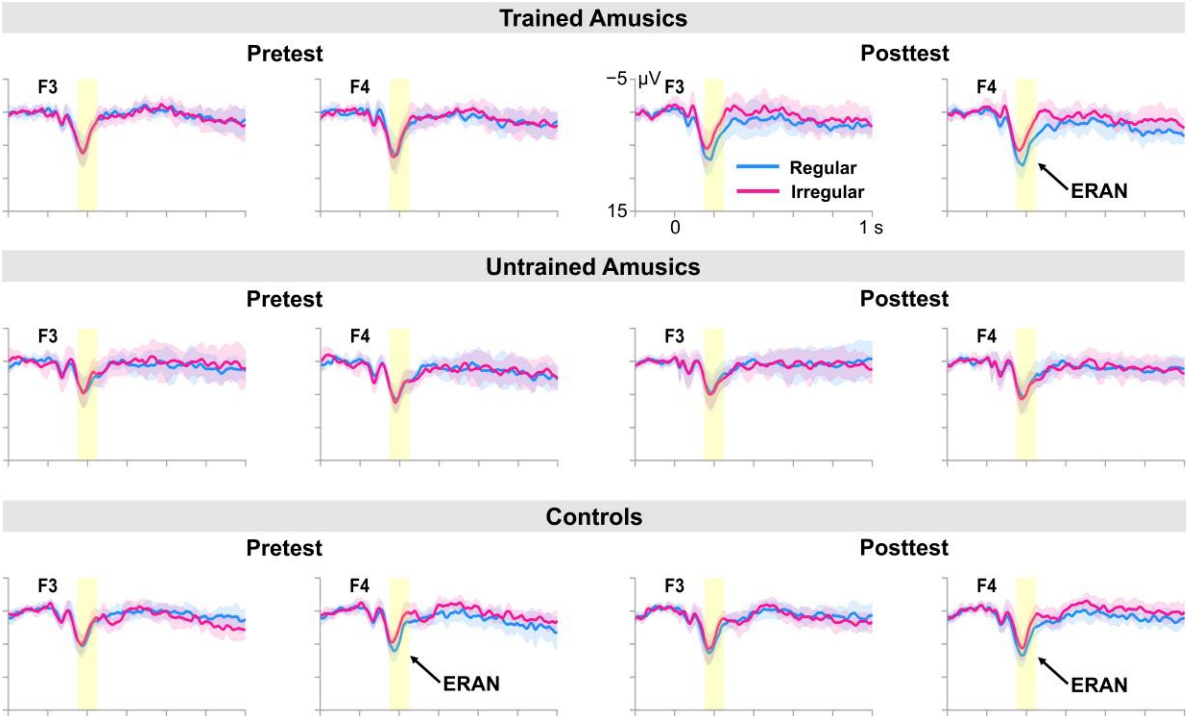
Expectedness Ratings by the Three Groups for Two Time Points



Note. Panel A: Raw ratings of regular and irregular melodies for each participant. Panel B: Difference scores between the raw ratings (irregular minus regular) for each participant. Each colored dot represents a single individual with lines connecting the same participant who has the same color in all plots. These dotplots were produced with ggplot2 (Version 3.4.0).

Figure 5

Grand Averaged ERPs at the Selected Scalp Sites as a Function of Time and Regularity



Note. Lines and shaded areas signify the average amplitudes and 95% CIs over subjects.

The time window of 150–250 ms is highlighted by light yellow bars.

Figure 6

Scalp Topographies of the ERAN (150–250 ms) for Each Group and Time Point

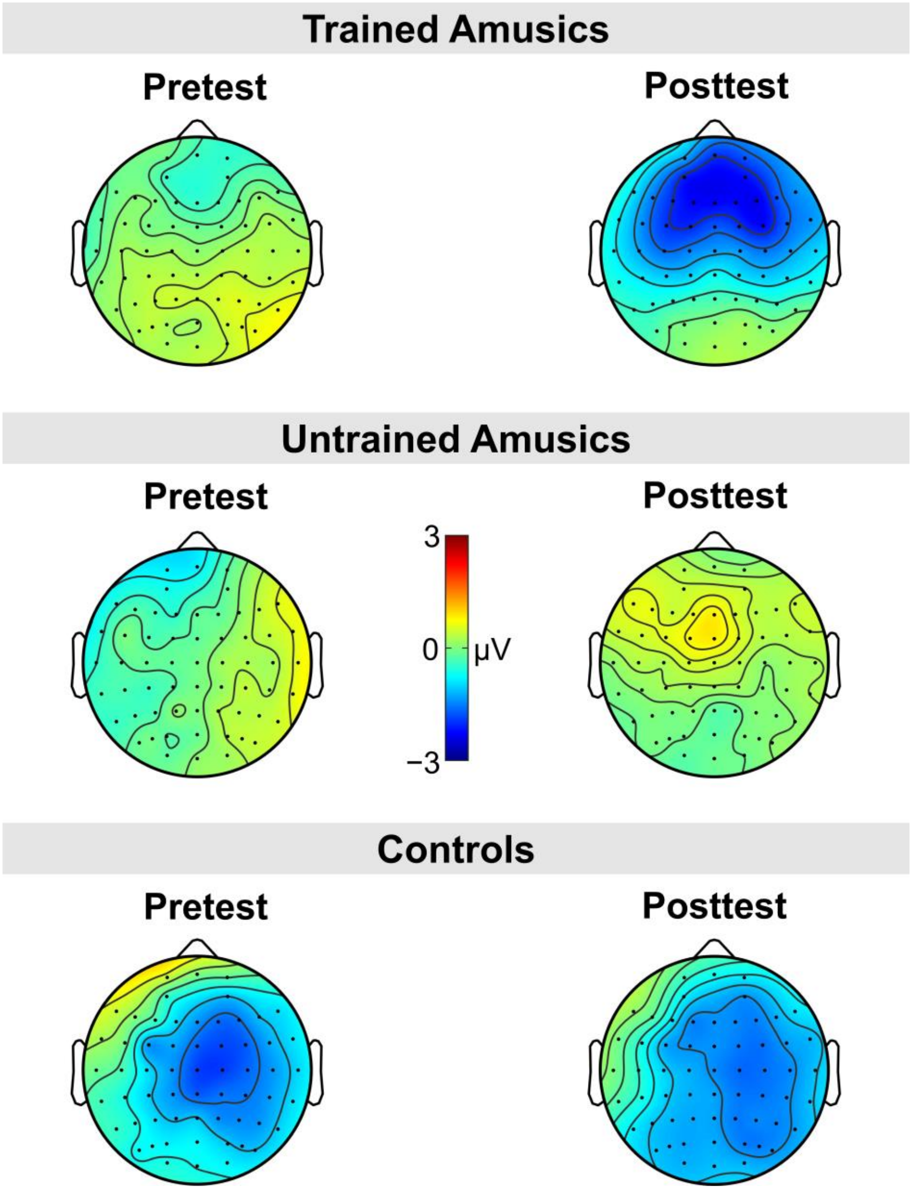
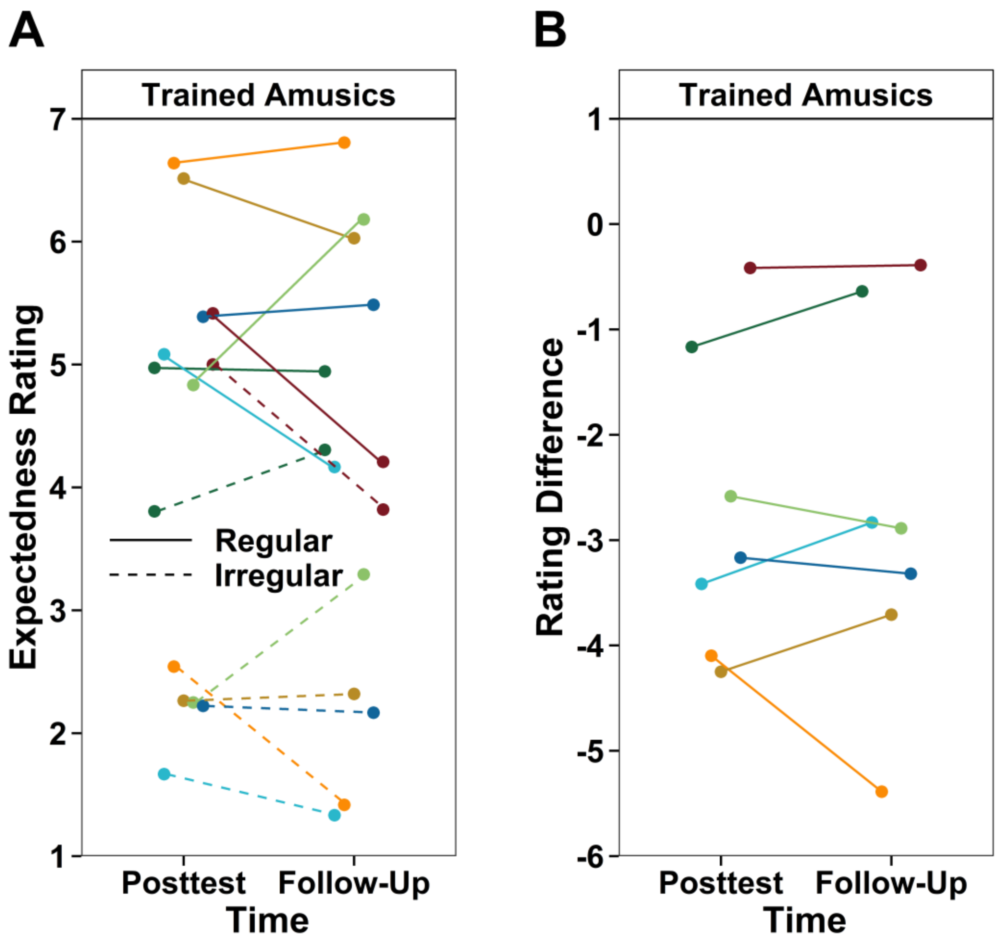


Figure 7

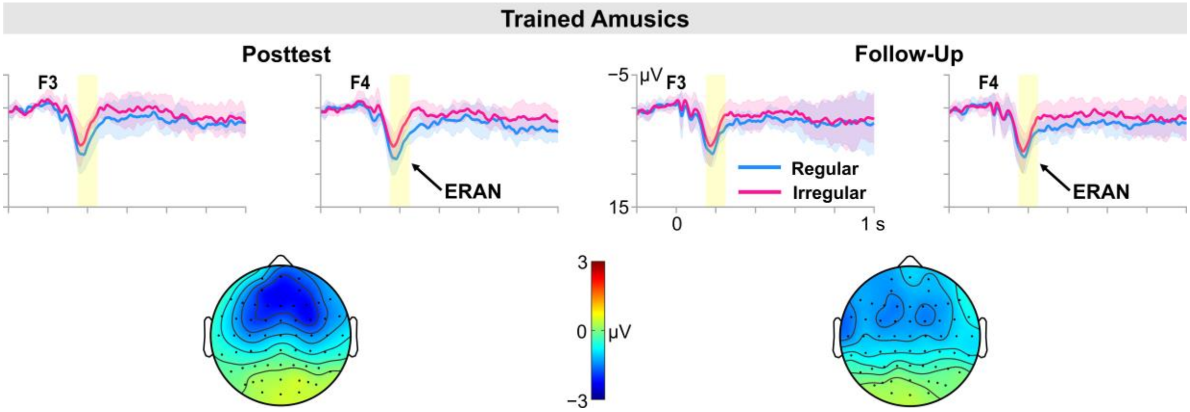
Expectedness Ratings by Trained Amusics at Each Time Point



Note. Panel A: Raw ratings of regular and irregular melodies for each participant. Panel B: Difference scores between the raw ratings (irregular minus regular) for each participant. Each colored dot represents a single individual with lines connecting the same participant who has the same color in all plots. These dotplots were produced with ggplot2 (Version 3.4.0).

Figure 8

Irregular Melodies Evoked an ERAN With a Bilateral Frontal Scalp Distribution (150–250 ms) for Trained Amusics at Each Time Point



Note. Lines and shaded areas denote the means ± 95% CIs. The time window of 150–250 ms is highlighted by light yellow bars.

Table S1 *Post Hoc Comparisons Among Groups in MBEA Measures and Pitch Threshold*

Measure	Comparison	Uncorrected *t* test	Corrected *t* test
MBEA			
Scale	T vs. U	$t(24) = 2.39, p = .025, d = 1.20$, 95% CI [0.10, 2.29]	$t(13.75) = 2.05, p = .136, d = 1.03$, 95% CI [−0.04, 2.06]
	C vs. T	$t(24) = 7.87, p < .001, d = 3.66$, 95% CI [2.20, 5.11]	$t(11.28) = 8.34, p < .001, d = 4.00$, 95% CI [2.03, 5.68]
	C vs. U	$t(24) = 10.44, p < .001, d = 4.85$, 95% CI [3.12, 6.59]	$t(10.31) = 9.95, p < .001, d = 4.80$, 95% CI [2.44, 6.77]
Contour	T vs. U	$t(24) = 0.00, p = 1.000, d = 0.00$, 95% CI [−1.03, 1.03]	$t(13.89) = 0.00, p = 1.000, d = 0.00$, 95% CI [−0.98, 0.98]
	C vs. T	$t(24) = 6.70, p < .001, d = 3.11$, 95% CI [1.78, 4.45]	$t(8.19) = 7.03, p < .001, d = 3.46$, 95% CI [1.44, 5.04]
	C vs. U	$t(24) = 6.70, p < .001, d = 3.11$, 95% CI [1.78, 4.45]	$t(7.99) = 6.46, p < .001, d = 3.19$, 95% CI [1.27, 4.68]
Interval	T vs. U	$t(24) = 0.44, p = .666, d = 0.22$, 95% CI [−0.82, 1.25]	$t(12.99) = 0.36, p = .933, d = 0.18$, 95% CI [−0.81, 1.16]
	C vs. T	$t(24) = 7.11, p < .001, d = 3.30$, 95% CI [1.93, 4.68]	$t(9.85) = 8.19, p < .001, d = 3.97$, 95% CI [1.90, 5.67]
	C vs. U	$t(24) = 7.58, p < .001, d = 3.52$, 95% CI [2.10, 4.94]	$t(8.62) = 6.82, p < .001, d = 3.34$, 95% CI [1.42, 4.87]
Rhythm	T vs. U	$t(24) = 0.68, p = .503, d = 0.34$, 95% CI [−0.70, 1.38]	$t(13.93) = 0.56, p = .845, d = 0.28$, 95% CI [−0.71, 1.26]
	C vs. T	$t(24) = 3.87, p < .001, d = 1.80$, 95% CI [0.70, 2.90]	$t(9.21) = 4.01, p = .007, d = 1.96$, 95% CI [0.60, 3.08]
	C vs. U	$t(24) = 4.61, p < .001, d = 2.14$, 95% CI [0.99, 3.29]	$t(8.92) = 4.48, p = .004, d = 2.19$, 95% CI [0.74, 3.37]
Meter	T vs. U	$t(24) = −1.64, p = .115, d = −0.82$, 95% CI [−1.88, 0.24]	$t(10.75) = −1.33, p = .408, d = −0.67$, 95% CI [−1.67, 0.37]
	C vs. T	$t(24) = 6.29, p < .001, d = 2.92$, 95% CI [1.63, 4.22]	$t(11.26) = 8.77, p < .001, d = 4.20$, 95% CI [2.16, 5.95]

	C vs. U	$t(24) = 4.53, p < .001, d = 2.10$, 95% CI [0.96, 3.25]	$t(8.27) = 3.73, p = .013, d = 1.84$, 95% CI [0.48, 2.93]
Memory	T vs. U	$t(24) = 0.20, p = .840, d = 0.10$, 95% CI [−0.93, 1.13]	$t(12.22) = 0.16, p = .985, d = 0.08$, 95% CI [−0.90, 1.06]
	C vs. T	$t(24) = 4.13, p < .001, d = 1.92$, 95% CI [0.80, 3.04]	$t(9.84) = 5.10, p = .001, d = 2.47$, 95% CI [0.97, 3.72]
	C vs. U	$t(24) = 4.35, p < .001, d = 2.02$, 95% CI [0.89, 3.16]	$t(8.28) = 3.77, p = .013, d = 1.86$, 95% CI [0.49, 2.96]
Melodic score	T vs. U	$t(24) = 1.19, p = .245, d = 0.60$, 95% CI [−0.45, 1.64]	$t(13.04) = 0.98, p = .601, d = 0.49$, 95% CI [−0.52, 1.48]
	C vs. T	$t(24) = 10.12, p < .001, d = 4.70$, 95% CI [3.01, 6.40]	$t(10.21) = 11.58, p < .001, d = 5.60$, 95% CI [2.90, 7.84]
	C vs. U	$t(24) = 11.40, p < .001, d = 5.30$, 95% CI [3.45, 7.15]	$t(8.86) = 10.30, p < .001, d = 5.04$, 95% CI [2.40, 7.14]
Global score (%)	T vs. U	$t(24) = 0.41, p = .686, d = 0.20$, 95% CI [−0.83, 1.24]	$t(13.95) = 0.33, p = .942, d = 0.16$, 95% CI [−0.82, 1.14]
	C vs. T	$t(24) = 11.77, p < .001, d = 5.47$, 95% CI [3.58, 7.36]	$t(8.61) = 12.13, p < .001, d = 5.95$, 95% CI [2.86, 8.39]
	C vs. U	$t(24) = 12.21, p < .001, d = 5.67$, 95% CI [3.73, 7.61]	$t(8.43) = 11.93, p < .001, d = 5.86$, 95% CI [2.78, 8.28]
Pitch perception threshold (semitones)			
Pitch direction discrimination	T vs. U	$t(24) = 0.99, p = .332, d = 0.49$, 95% CI [−0.55, 1.54]	$t(13.60) = 1.11, p = .524, d = 0.56$, 95% CI [−0.46, 1.55]
	C vs. T	$t(24) = -3.81, p < .001, d = -1.77$, 95% CI [−2.87, −0.68]	$t(16.49) = -3.67, p = .005, d = -1.68$, 95% CI [−2.76, −0.61]
	C vs. U	$t(24) = -2.75, p = .011, d = -1.28$, 95% CI [−2.31, −0.25]	$t(17.00) = -2.86, p = .028, d = -1.29$, 95% CI [−2.33, −0.30]

Note. T = trained amusics; U = untrained amusics; C = controls. Effect sizes and confidence intervals in uncorrected and corrected t tests were obtained using jamovi (Version 2.3.21) and the R package misty (Version 0.4.6), respectively.

Table S2

Table S2 *Post Hoc Comparisons of Nine Training Sessions*

Comparison	$t(7)$	$p_{uncorrected}$	p_{holm}	d	95% CI
Session 2 vs. Session 1	2.81	.026	.736	1.06	[0.12, 1.96]
Session 3 vs. Session 1	2.76	.028	.739	1.04	[0.11, 1.93]
Session 4 vs. Session 1	3.96	.005	.180	1.50	[0.41, 2.54]
Session 5 vs. Session 1	4.00	.005	.176	1.51	[0.42, 2.56]
Session 6 vs. Session 1	2.78	.027	.739	1.05	[0.11, 1.94]
Session 7 vs. Session 1	3.19	.015	.442	1.21	[0.22, 2.15]
Session 8 vs. Session 1	6.08	< .001	.018	2.30	[0.89, 3.67]
Session 9 vs. Session 1	4.96	.002	.057	1.87	[0.64, 3.07]
Session 3 vs. Session 2	0.13	.903	1.000	0.05	[−0.69, 0.79]
Session 4 vs. Session 2	1.84	.109	1.000	0.70	[−0.15, 1.50]
Session 5 vs. Session 2	2.16	.068	1.000	0.82	[−0.06, 1.65]
Session 6 vs. Session 2	1.15	.288	1.000	0.43	[−0.35, 1.19]
Session 7 vs. Session 2	1.84	.108	1.000	0.70	[−0.15, 1.50]
Session 8 vs. Session 2	3.23	.014	.433	1.22	[0.23, 2.17]
Session 9 vs. Session 2	2.32	.053	1.000	0.88	[−0.01, 1.72]
Session 4 vs. Session 3	1.98	.089	1.000	0.75	[−0.11, 1.56]
Session 5 vs. Session 3	1.40	.204	1.000	0.53	[−0.28, 1.30]
Session 6 vs. Session 3	0.93	.385	1.000	0.35	[−0.42, 1.10]
Session 7 vs. Session 3	2.26	.059	1.000	0.85	[−0.03, 1.69]
Session 8 vs. Session 3	3.86	.006	.198	1.46	[0.38, 2.49]
Session 9 vs. Session 3	3.26	.014	.429	1.23	[0.23, 2.18]
Session 5 vs. Session 4	0.49	.640	1.000	0.19	[−0.57, 0.93]
Session 6 vs. Session 4	−0.10	.920	1.000	−0.04	[−0.78, 0.70]
Session 7 vs. Session 4	1.05	.328	1.000	0.40	[−0.38, 1.15]
Session 8 vs. Session 4	2.34	.052	1.000	0.88	[−0.01, 1.73]
Session 9 vs. Session 4	1.87	.103	1.000	0.71	[−0.14, 1.51]
Session 6 vs. Session 5	−0.35	.737	1.000	−0.13	[−0.87, 0.62]

Session 7 vs. Session 5	0.58	.581	1.000	0.22	[−0.54, 0.96]
Session 8 vs. Session 5	2.38	.049	1.000	0.90	[0.00, 1.75]
Session 9 vs. Session 5	1.57	.160	1.000	0.59	[−0.23, 1.38]
Session 7 vs. Session 6	1.11	.305	1.000	0.42	[−0.37, 1.18]
Session 8 vs. Session 6	1.96	.090	1.000	0.74	[−0.11, 1.55]
Session 9 vs. Session 6	1.60	.154	1.000	0.60	[−0.22, 1.39]
Session 8 vs. Session 7	1.33	.224	1.000	0.50	[−0.30, 1.27]
Session 9 vs. Session 7	0.66	.531	1.000	0.25	[−0.51, 0.99]
Session 9 vs. Session 8	−0.45	.666	1.000	−0.17	[−0.91, 0.58]

Table S3 *Post Hoc Comparisons for the Time × Group × Regularity Interaction in Behavioral Data*

Comparison	*t*	*df*	$p_{uncorrected}$	p_{holm}	*d*	95% CI
Pre-T-R vs. Pre-T-I	1.89	72	.063	1.000	0.22	[−0.01, 0.46]
Pre-T-R vs. Pre-U-R	1.77	93.28	.080	1.000	0.37	[−0.04, 0.77]
Pre-T-R vs. Pre-U-I	3.43	93.28	.001	.030	0.71	[0.29, 1.13]
Pre-T-I vs. Pre-U-R	−0.02	93.28	.984	1.000	−0.004	[−0.41, 0.40]
Pre-T-I vs. Pre-U-I	1.64	93.28	.105	1.000	0.34	[−0.07, 0.75]
Pre-U-R vs. Pre-U-I	1.74	72	.085	1.000	0.21	[−0.03, 0.44]
Pre-C-R vs. Pre-C-I	8.39	72	< .001	< .001	0.99	[0.70, 1.27]
Pre-C-R vs. Pre-T-R	1.55	93.28	.126	1.000	0.32	[−0.09, 0.73]
Pre-C-R vs. Pre-T-I	3.47	93.28	.001	.028	0.72	[0.30, 1.14]
Pre-C-R vs. Pre-U-R	3.45	93.28	.001	.029	0.71	[0.29, 1.13]
Pre-C-R vs. Pre-U-I	5.23	93.28	< .001	< .001	1.08	[0.65, 1.52]
Pre-C-I vs. Pre-T-R	−5.77	93.28	< .001	< .001	−1.19	[−1.63, −0.75]
Pre-C-I vs. Pre-T-I	−3.84	93.28	< .001	.009	−0.80	[−1.21, −0.37]
Pre-C-I vs. Pre-U-R	−3.86	93.28	< .001	.009	−0.80	[−1.22, −0.38]
Pre-C-I vs. Pre-U-I	−2.08	93.28	.040	.925	−0.43	[−0.84, −0.02]
Post-T-R vs. Post-T-I	8.18	72	< .001	< .001	0.96	[0.68, 1.24]
Post-T-R vs. Post-U-R	2.46	93.28	.016	.412	0.51	[0.10, 0.92]
Post-T-R vs. Post-U-I	3.99	93.28	< .001	.006	0.83	[0.40, 1.25]
Post-T-R vs. Pre-C-R	0.43	93.28	.666	1.000	0.09	[−0.32, 0.49]
Post-T-R vs. Pre-C-I	7.75	93.28	< .001	< .001	1.60	[1.14, 2.07]
Post-T-R vs. Pre-T-R	1.94	72	.057	1.000	0.23	[−0.01, 0.46]
Post-T-R vs. Pre-T-I	3.82	72	< .001	.011	0.45	[0.21, 0.69]
Post-T-R vs. Pre-U-R	3.61	93.28	< .001	.018	0.75	[0.33, 1.17]
Post-T-R vs. Pre-U-I	5.27	93.28	< .001	< .001	1.09	[0.65, 1.52]
Post-T-I vs. Post-U-R	−5.31	93.28	< .001	< .001	−1.10	[−1.53, −0.66]
Post-T-I vs. Post-U-I	−3.77	93.28	< .001	.011	−0.78	[−1.20, −0.36]

Post-T-I vs. Pre-C-R	−7.92	93.28	< .001	< .001	−1.64	[−2.11, −1.17]
Post-T-I vs. Pre-C-I	−0.61	93.28	.545	1.000	−0.13	[−0.53, 0.28]
Post-T-I vs. Pre-T-R	−6.24	72	< .001	< .001	−0.74	[−0.99, −0.47]
Post-T-I vs. Pre-T-I	−4.35	72	< .001	.002	−0.51	[−0.76, −0.27]
Post-T-I vs. Pre-U-R	−4.15	93.28	< .001	.003	−0.86	[−1.28, −0.43]
Post-T-I vs. Pre-U-I	−2.50	93.28	.014	.385	−0.52	[−0.93, −0.10]
Post-U-R vs. Post-U-I	1.62	72	.110	1.000	0.19	[−0.04, 0.42]
Post-U-R vs. Pre-C-R	−2.21	93.28	.029	.736	−0.46	[−0.87, −0.05]
Post-U-R vs. Pre-C-I	5.10	93.28	< .001	< .001	1.06	[0.62, 1.49]
Post-U-R vs. Pre-T-R	−0.62	93.28	.537	1.000	−0.13	[−0.53, 0.28]
Post-U-R vs. Pre-T-I	1.17	93.28	.244	1.000	0.24	[−0.17, 0.65]
Post-U-R vs. Pre-U-R	1.21	72	.229	1.000	0.14	[−0.09, 0.37]
Post-U-R vs. Pre-U-I	2.96	72	.004	.130	0.35	[0.11, 0.59]
Post-U-I vs. Pre-C-R	−3.86	93.28	< .001	.009	−0.80	[−1.22, −0.38]
Post-U-I vs. Pre-C-I	3.45	93.28	.001	.029	0.71	[0.29, 1.13]
Post-U-I vs. Pre-T-R	−2.15	93.28	.034	.811	−0.45	[−0.85, −0.03]
Post-U-I vs. Pre-T-I	−0.36	93.28	.717	1.000	−0.07	[−0.48, 0.33]
Post-U-I vs. Pre-U-R	−0.40	72	.687	1.000	−0.05	[−0.28, 0.18]
Post-U-I vs. Pre-U-I	1.34	72	.184	1.000	0.16	[−0.08, 0.39]
Post-C-R vs. Post-C-I	8.61	72	< .001	< .001	1.01	[0.73, 1.30]
Post-C-R vs. Post-T-R	−1.01	93.28	.314	1.000	−0.21	[−0.62, 0.20]
Post-C-R vs. Post-T-I	7.34	93.28	< .001	< .001	1.52	[1.06, 1.98]
Post-C-R vs. Post-U-R	1.63	93.28	.106	1.000	0.34	[−0.07, 0.75]
Post-C-R vs. Post-U-I	3.28	93.28	.001	.046	0.68	[0.26, 1.09]
Post-C-R vs. Pre-C-R	−0.67	72	.508	1.000	−0.08	[−0.31, 0.15]
Post-C-R vs. Pre-C-I	7.73	72	< .001	< .001	0.91	[0.63, 1.18]
Post-C-R vs. Pre-T-R	0.97	93.28	.337	1.000	0.20	[−0.21, 0.61]
Post-C-R vs. Pre-T-I	2.89	93.28	.005	.143	0.60	[0.18, 1.01]
Post-C-R vs. Pre-U-R	2.87	93.28	.005	.147	0.59	[0.18, 1.01]

Post-C-R vs. Pre-U-I	4.65	93.28	< .001	.001	0.96	[0.53, 1.39]
Post-C-I vs. Post-T-R	−8.51	93.28	< .001	< .001	−1.76	[−2.24, −1.28]
Post-C-I vs. Post-T-I	−0.16	93.28	.876	1.000	−0.03	[−0.44, 0.37]
Post-C-I vs. Post-U-R	−5.87	93.28	< .001	< .001	−1.22	[−1.65, −0.77]
Post-C-I vs. Post-U-I	−4.21	93.28	< .001	.003	−0.87	[−1.29, −0.45]
Post-C-I vs. Pre-C-R	−9.27	72	< .001	< .001	−1.09	[−1.38, −0.80]
Post-C-I vs. Pre-C-I	−0.88	72	.383	1.000	−0.10	[−0.33, 0.13]
Post-C-I vs. Pre-T-R	−6.53	93.28	< .001	< .001	−1.35	[−1.80, −0.90]
Post-C-I vs. Pre-T-I	−4.61	93.28	< .001	.001	−0.95	[−1.38, −0.52]
Post-C-I vs. Pre-U-R	−4.63	93.28	< .001	.001	−0.96	[−1.38, −0.53]
Post-C-I vs. Pre-U-I	−2.84	93.28	.005	.153	−0.59	[−1.00, −0.17]

Note. Pre = pretest; Post = posttest; T = trained amusics; U = untrained amusics; C = controls; R = regular, I = irregular.

Table S4 *Post Hoc Comparisons for the Time × Group × Regularity Interaction in ERP Data*

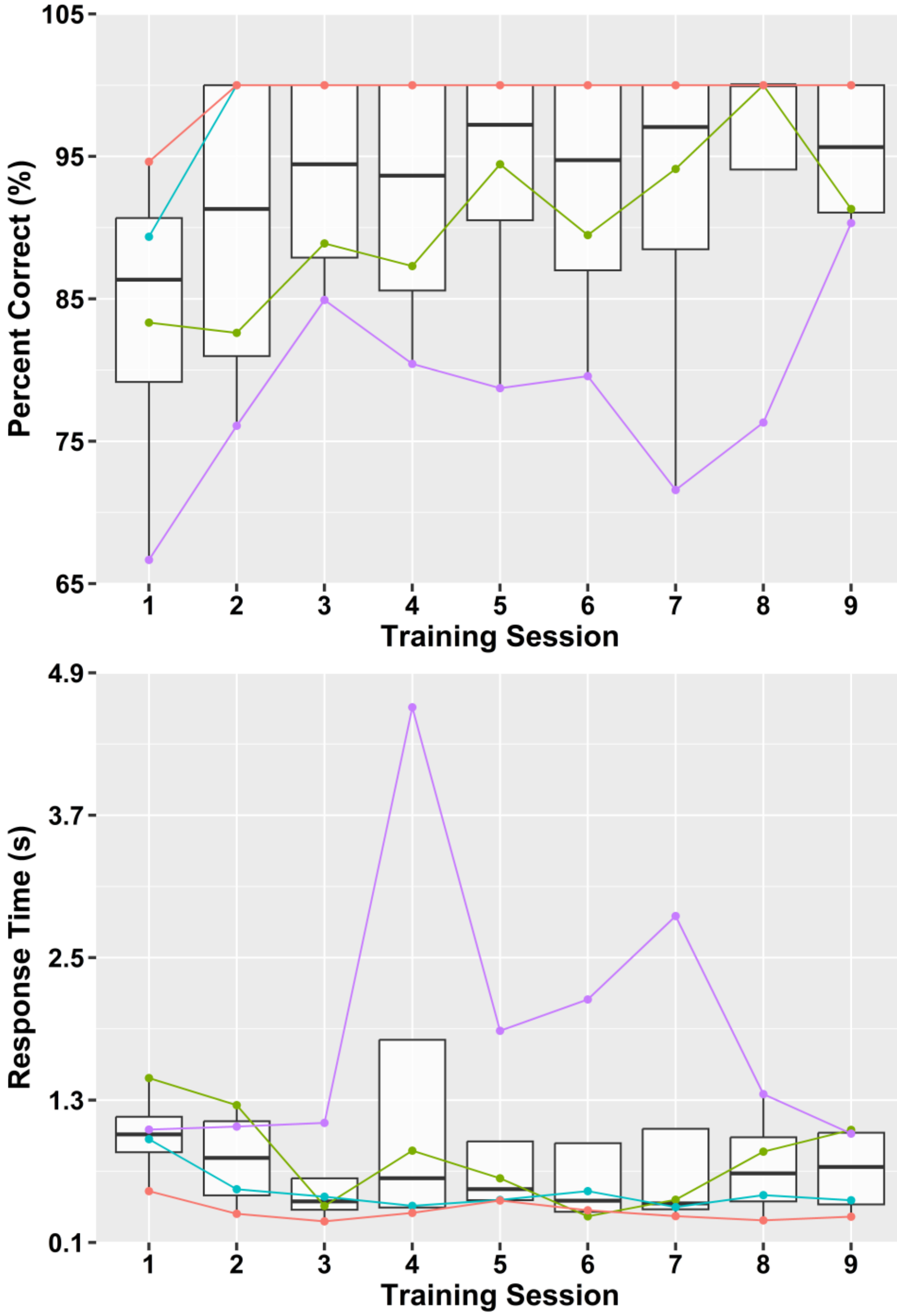
Comparison	$t(24)$	$p_{uncorrected}$	p_{holm}	d	95% CI
Pre-T-R vs. Pre-T-I	0.26	.799	1.000	0.05	[−0.35, 0.45]
Pre-T-R vs. Pre-U-R	0.50	.624	1.000	0.20	[−0.60, 1.00]
Pre-T-R vs. Pre-U-I	0.70	.492	1.000	0.29	[−0.52, 1.09]
Pre-T-I vs. Pre-U-R	0.42	.680	1.000	0.17	[−0.63, 0.97]
Pre-T-I vs. Pre-U-I	0.64	.530	1.000	0.26	[−0.54, 1.06]
Pre-U-R vs. Pre-U-I	0.32	.753	1.000	0.07	[−0.34, 0.47]
Pre-C-R vs. Pre-C-I	3.39	.002	.149	0.69	[0.24, 1.13]
Pre-C-R vs. Pre-T-R	−0.61	.547	1.000	−0.25	[−1.05, 0.56]
Pre-C-R vs. Pre-T-I	−0.54	.593	1.000	−0.22	[−1.02, 0.58]
Pre-C-R vs. Pre-U-R	−0.08	.940	1.000	−0.03	[−0.83, 0.77]
Pre-C-R vs. Pre-U-I	0.08	.934	1.000	0.03	[−0.77, 0.83]
Pre-C-I vs. Pre-T-R	−2.14	.043	1.000	−0.87	[−1.70, −0.03]
Pre-C-I vs. Pre-T-I	−2.27	.033	1.000	−0.93	[−1.76, −0.08]
Pre-C-I vs. Pre-U-R	−1.56	.131	1.000	−0.64	[−1.45, 0.19]
Pre-C-I vs. Pre-U-I	−1.58	.127	1.000	−0.65	[−1.46, 0.18]
Post-T-R vs. Post-T-I	4.49	< .001	.010	0.92	[0.43, 1.39]
Post-T-R vs. Post-U-R	2.12	.045	1.000	0.87	[0.02, 1.69]
Post-T-R vs. Post-U-I	1.75	.093	1.000	0.71	[−0.12, 1.53]
Post-T-R vs. Pre-C-R	2.25	.034	1.000	0.92	[0.07, 1.75]
Post-T-R vs. Pre-C-I	3.65	.001	.078	1.49	[0.57, 2.38]
Post-T-R vs. Pre-T-R	3.67	.001	.076	0.75	[0.29, 1.20]
Post-T-R vs. Pre-T-I	3.84	< .001	.050	0.78	[0.32, 1.23]
Post-T-R vs. Pre-U-R	2.06	.051	1.000	0.84	[0.00, 1.67]
Post-T-R vs. Pre-U-I	2.34	.028	1.000	0.96	[0.10, 1.79]
Post-T-I vs. Post-U-R	0.10	.918	1.000	0.04	[−0.76, 0.84]
Post-T-I vs. Post-U-I	−0.45	.654	1.000	−0.18	[−0.98, 0.62]

Post-T-I vs. Pre-C-R	−0.11	.910	1.000	−0.04	[−0.84, 0.76]
Post-T-I vs. Pre-C-I	1.30	.205	1.000	0.53	[−0.29, 1.34]
Post-T-I vs. Pre-T-R	−1.10	.282	1.000	−0.22	[−0.63, 0.18]
Post-T-I vs. Pre-T-I	−1.18	.250	1.000	−0.24	[−0.64, 0.17]
Post-T-I vs. Pre-U-R	−0.18	.862	1.000	−0.07	[−0.87, 0.73]
Post-T-I vs. Pre-U-I	−0.04	.967	1.000	−0.02	[−0.82, 0.78]
Post-U-R vs. Post-U-I	−1.08	.289	1.000	−0.22	[−0.62, 0.19]
Post-U-R vs. Pre-C-R	−0.22	.830	1.000	−0.09	[−0.89, 0.71]
Post-U-R vs. Pre-C-I	1.05	.304	1.000	0.43	[−0.38, 1.23]
Post-U-R vs. Pre-T-R	−0.71	.484	1.000	−0.29	[−1.09, 0.52]
Post-U-R vs. Pre-T-I	−0.65	.520	1.000	−0.27	[−1.07, 0.54]
Post-U-R vs. Pre-U-R	−0.61	.549	1.000	−0.12	[−0.52, 0.28]
Post-U-R vs. Pre-U-I	−0.32	.754	1.000	−0.07	[−0.47, 0.34]
Post-U-I vs. Pre-C-R	0.39	.703	1.000	0.16	[−0.64, 0.96]
Post-U-I vs. Pre-C-I	1.84	.079	1.000	0.75	[−0.08, 1.57]
Post-U-I vs. Pre-T-R	−0.19	.850	1.000	−0.08	[−0.88, 0.72]
Post-U-I vs. Pre-T-I	−0.09	.932	1.000	−0.04	[−0.84, 0.76]
Post-U-I vs. Pre-U-R	0.49	.630	1.000	0.10	[−0.30, 0.50]
Post-U-I vs. Pre-U-I	0.92	.364	1.000	0.19	[−0.22, 0.59]
Post-C-R vs. Post-C-I	3.17	.004	.245	0.65	[0.20, 1.08]
Post-C-R vs. Post-T-R	−1.09	.287	1.000	−0.44	[−1.25, 0.37]
Post-C-R vs. Post-T-I	1.18	.250	1.000	0.48	[−0.33, 1.29]
Post-C-R vs. Post-U-R	1.19	.245	1.000	0.49	[−0.33, 1.29]
Post-C-R vs. Post-U-I	0.72	.477	1.000	0.29	[−0.51, 1.10]
Post-C-R vs. Pre-C-R	2.69	.013	0.750	0.55	[0.12, 0.97]
Post-C-R vs. Pre-C-I	5.65	< .001	< .001	1.15	[0.63, 1.66]
Post-C-R vs. Pre-T-R	0.55	.586	1.000	0.22	[−0.58, 1.02]
Post-C-R vs. Pre-T-I	0.72	.478	1.000	0.29	[−0.51, 1.10]
Post-C-R vs. Pre-U-R	1.04	.310	1.000	0.42	[−0.39, 1.23]

Post-C-R vs. Pre-U-I	1.28	.214	1.000	0.52	[−0.30, 1.33]
Post-C-I vs. Post-T-R	−2.54	.018	1.000	−1.04	[−1.88, −0.18]
Post-C-I vs. Post-T-I	−0.26	.798	1.000	−0.11	[−0.91, 0.70]
Post-C-I vs. Post-U-R	−0.13	.901	1.000	−0.05	[−0.85, 0.75]
Post-C-I vs. Post-U-I	−0.75	.462	1.000	−0.31	[−1.11, 0.50]
Post-C-I vs. Pre-C-R	−0.70	.493	1.000	−0.14	[−0.54, 0.26]
Post-C-I vs. Pre-C-I	2.25	.034	1.000	0.46	[0.03, 0.88]
Post-C-I vs. Pre-T-R	−0.98	.337	1.000	−0.40	[−1.20, 0.41]
Post-C-I vs. Pre-T-I	−0.95	.353	1.000	−0.39	[−1.19, 0.42]
Post-C-I vs. Pre-U-R	−0.46	.651	1.000	−0.19	[−0.99, 0.62]
Post-C-I vs. Pre-U-I	−0.34	.736	1.000	−0.14	[−0.94, 0.66]

Note. Pre = pretest; Post = posttest; T = trained amusics; U = untrained amusics; C = controls; R = regular, I = irregular.

Figure S1 *Percentage and Response Time of Correct Responses Across the Nine Training Sessions for Trained Controls*



Note. Trials were excluded from the analysis if the reaction time was shorter than 200 ms or greater than 3 *SD* above the individual's mean for every training session. Each data point represents an individual sample and the two same color lines belong to the same subject.

These data are summarized in boxplots in which the bisecting line in the box plot represents the median, the box represents the 25th and 75th percentiles, and the whisker represents 1.5 times the interquartile range from the 25th and 75th percentiles. The graphs were maded with the GGally (Version 2.1.2) and ggplot2 (Version 3.4.0) packages within the R.

Author contributions

Jun Jiang: Conceptualization, Methodology, Investigation, Resources, Software, Formal analysis, Data curation, Visualization, Validation, Writing–original draft, Writing–review & editing. **Fang Liu:** Conceptualization, Funding acquisition, Writing–review & editing. **Linshu Zhou:** Methodology, Investigation, Funding acquisition. **Liaoliao Chen:** Investigation, Resources. **Cunmei Jiang:** Conceptualization, Methodology, Resources, Supervision, Project administration, Funding acquisition, Writing–review & editing.