



# Musical ability and emotion recognition in speech prosody: The role of pitch discrimination

Aíssa M. Baldé<sup>1</sup> · E. Glenn Schellenberg<sup>1,2</sup> · César F. Lima<sup>1</sup>

Received: 25 November 2025 / Accepted: 17 January 2026  
© The Author(s) 2026

## Abstract

Why does musical expertise predict enhanced emotion recognition in speech prosody? Evidence for a causal role of music training is weak, and correlations with musical aptitude could reflect basic auditory abilities rather than musicality per se. Here, we tested whether individual differences in basic auditory discrimination account for the music–prosody association. A total of 164 adults completed forced-choice judgments of emotions in prosody and facial expressions, self-reports of musical experience, objective tests of musical ability (melody and rhythm perception), and adaptive psychoacoustic tasks that estimated discrimination thresholds for pitch, duration, loudness, timbre, and backward masking. Both music training and musical ability correlated with better recognition of prosodic but not facial emotions. The training effect was weak, however, and disappeared after controlling for confounding variables, including general cognitive ability. By contrast, musical ability, specifically melody perception, remained associated with prosodic emotion recognition after accounting for training and other covariates. Crucially, psychoacoustic thresholds correlated with both prosody recognition and melody perception. When considered simultaneously, pitch discrimination alone independently predicted prosodic emotion recognition, but melody perception did not. These findings suggest that music training is an artifactual correlate of prosodic emotion recognition, and that basic pitch sensitivity underlies the link between musical ability and emotional prosody.

**Keywords** Emotion recognition · Music training · Musical aptitude · Psychoacoustics · Pitch discrimination

Recognizing emotions from speech prosody, through intonation, timing, loudness, and timbre, is essential for communication and socioemotional functioning (Carton et al., 1999; Leipold et al., 2022; Neves et al., 2021; Paz et al., 2025; Terracciano et al., 2003). Listeners reliably identify prosodic emotions from early in development (Sauter et al., 2013) and across cultures (Cowen et al., 2019; Scherer et al., 2001). Emotions in music show a remarkably similar profile. They are recognized from early childhood (Vidas et al., 2018), cross-culturally (Fritz et al., 2009; Singh & Mehr, 2023), and

through acoustic cues that largely overlap with those used in prosody (Juslin & Laukka, 2003). Sadness, for example, is typically conveyed by slower tempo or speech rate, lower sound level or vocal intensity, lower pitch, and reduced high-frequency energy. These parallels raise the possibility that sensitivity to emotional cues generalizes across domains.

A growing literature reports positive correlations between music training and emotion recognition, both in music (Bhatara et al., 2011; Castro & Lima, 2014) and in prosody (Correia et al., 2022a, 2022b; Jansen et al., 2023; Lima & Castro, 2011; Martins et al., 2021; Nussbaum et al., 2024). Although associations with prosody are typically small and not always replicable (Martins et al., 2021; Nussbaum et al., 2025), they are theoretically important as potential evidence of transfer from music training to nonmusical communication. The OPERA hypothesis (Patel, 2011, 2014) formalizes one plausible mechanism: Because music and speech share sensory and cognitive processes, music training could enhance speech processing by involving greater sensory precision, extensive practice, focused attention, and emotional engagement.

---

E. Glenn Schellenberg and César F. Lima contributed equally to this work.

✉ César F. Lima  
cesar.lima@iscte-iul.pt

<sup>1</sup> Centro de Investigação e Intervenção Social (CIS-IUL), Instituto Universitário de Lisboa (ISCTE-IUL), Av.<sup>a</sup> das Forças Armadas, 1649-026 Lisbon, Portugal

<sup>2</sup> Department of Psychology, University of Toronto Mississauga, Mississauga, Canada

Nevertheless, causal evidence for training effects is limited. Most studies are correlational, comparing trained and untrained individuals, which leaves open roles for selection effects and preexisting differences (Martins et al., 2021; Schellenberg, 2020; Schellenberg & Lima, 2024). Because music training is associated with socioeconomic status, general cognitive ability, and personality (e.g., Corrigan et al., 2013), any of these factors could explain observed advantages. Indeed, a recent review questioned whether music training improves nonmusical abilities, including emotion recognition (Schellenberg & Lima, 2024), and a large multi-lab study failed to replicate previously reported links with neural responses to speech (Whiteford et al., 2025).

An alternative account shifts the focus from training to musical aptitude or ability. Perceptual measures of musical ability often outperform years of lessons in predicting vocal-emotion recognition. For example, a randomized intervention found no improvement in children's emotion recognition after two years of music lessons, yet baseline musical ability predicted emotion recognition independently of training (Neves et al., 2025). Similarly, among adults, musical ability correlates with vocal-emotion recognition independently of training, whereas apparent training effects disappear after accounting for ability (Correia et al., 2022a, 2022b; Nussbaum et al., 2024, 2025). Meta-analytic evidence confirms that prosody perception, both emotional and linguistic, is more strongly associated with musical ability than with music training (Jansen et al., 2023).

Musical ability is commonly assessed with same-different judgments of short musical sequences that vary in melody, rhythm, or other features (e.g., Musical Ear Test [MET]; Wallentin et al., 2010; Profile of Music Perception Skills [PROMS]; Law & Zentner, 2012). Although typical sequences are structured musical patterns, music perception begins with encoding elementary acoustic features, such as pitch and duration, before music-specific operations occur (Koelsch, 2011; Peretz & Coltheart, 2003). Because stimuli typically comprise multiple tones, musical-ability tests also recruit short-term and working memory, which help to explain why musical-ability scores correlate with general cognitive performance (Baldé et al., 2025; Correia et al., 2022a, 2022b, 2023; Grassi et al., 2025; Swaminathan & Schellenberg, 2020).

In short, associations between musical ability and prosodic emotion recognition may reflect low-level auditory sensitivity or general cognitive mechanisms rather than “musicality” per se. While the speech and language literature routinely separates sensory encoding from higher-order linguistic processing (e.g., Saito, 2023), the music literature often treats psychoacoustic discrimination and complex music-perception tests interchangeably (Correia et al., 2022a, 2022b; Law & Zentner, 2012; Martin et al., 2018; Thompson et al., 2012). According to mechanistic models

of music perception and cognition (Koelsch, 2011; Peretz & Coltheart, 2003), domain-general auditory processes should be distinguished from music-specific operations.

Basic auditory processing is best measured with adaptive psychoacoustic tasks using simple, nonmusical stimuli (e.g., pure tones) to estimate sensory thresholds for isolated dimensions such as pitch, duration, intensity, and timbre (Baldé et al., 2025; Soranzo & Grassi, 2014). Lower (better) thresholds are typically evident in musicians across dimensions (Bianchi et al., 2016; Endo et al., 2021; Lee & Müllensiefen, 2020; Rammsayer & Altenmüller, 2006), whereas higher thresholds are evident among individuals with congenital amusia, a disorder characterized by impaired pitch discrimination (Hyde & Peretz, 2004). Moreover, psychoacoustic thresholds predict musical ability (melody and rhythm perception) even after controlling for training, cognitive abilities, personality, and demographics (Baldé et al., 2025).

Importantly, psychoacoustic thresholds—particularly pitch discrimination—predict prosodic emotion recognition in typically developing adults (Globerson et al., 2013) and in individuals with amusia (Lima et al., 2016; Lolli et al., 2015; Pralus et al., 2019). In a reanalysis of Correia et al., (2022a, 2022b; see the Supplementary Materials), psychoacoustic thresholds, but not a beat-alignment musical test, predicted vocal-emotion recognition independently of training. When considered simultaneously, only pitch discrimination accounted for unique variance in emotion recognition. Voice-morphing studies highlight further that when prosodic emotions are expressed exclusively through pitch, recognition remains high and correlated with musical expertise (Nussbaum et al., 2024).

Motivated by these findings, we asked whether musical expertise (defined as both training and ability) predicts emotion recognition from prosody, and whether any association is music-specific or explained by basic auditory sensitivity. We tested 164 adults with diverse musical backgrounds. Forced-choice tasks measured emotion recognition from prosody and, as a nonauditory control, facial expressions. Predictors included self-reported music training and abilities (Goldsmiths Musical Sophistication Index [Gold-MSI]; Müllensiefen et al., 2014), objective melody and rhythm perception (MET subtests; Wallentin et al., 2010), and psychoacoustic thresholds for pitch, duration, loudness, timbre, and backward masking. We also measured cognitive abilities, personality, and sociodemographic variables, which can covary with musical expertise. For example, training duration correlates with higher general cognitive ability and openness-to-experience (e.g., Corrigan et al., 2013), whereas musical ability improves with age, education, and general cognitive ability (e.g., Correia et al., 2022a, 2022b).

We hypothesized that (1) musical expertise—particularly musical ability—would correlate with prosodic emotion

recognition, (2) basic auditory processing would correlate with both musical ability and prosody perception, and (3) pitch discrimination would account for the association between musical ability and prosodic emotion recognition, such that musical ability would not explain unique variance after accounting for pitch thresholds.

## Method

### Participants

The initial sample comprised 167 participants but three were excluded due to missing data ( $n = 1$ ) or hearing difficulties ( $n = 2$ ). The final sample of 164 participants (138 women, 26 men) ranged in age from 18 to 65 years ( $M = 22.7$ ,  $SD = 7.28$ ). Age was square-root transformed for statistical analysis. Most participants were undergraduate students. Education was coded as a three-level ordinal variable: no bachelor's degree ( $n = 123$ ), bachelor's degree ( $n = 26$ ), or master's degree ( $n = 15$ ). All participants were native speakers of European Portuguese and reported normal hearing.

Participants varied widely in formal music training. Sixty-four reported no lessons, six reported half a year, nine reported 1 year, 17 reported 2 years, and 18 reported between 3 and 5 years. Fifty participants had 6 or more years of training, including 26 with at least 10 years.

Sample size was similar to that of Baldé et al. (2025) and Correia et al., (2022a, 2022b). Post hoc power analysis conducted with G\*Power (Faul et al., 2009) confirmed that we had 90.1% power to detect medium-sized partial correlations ( $r = .25$ , seven covariates, two-tailed,  $\alpha = .05$ ).

### Materials

#### Emotion recognition

Two forced-choice tasks required participants to categorize emotions conveyed by prosodic cues or facial expressions. Both tasks had 84 trials. The stimuli were taken from validated corpora (prosody, Castro & Lima, 2010; faces, Karolinska Directed Emotional Faces database, Lundqvist et al., 1998) and expressed one of seven emotions: anger, disgust, fear, happiness, surprise, sadness, and neutrality, with 12 stimuli per emotion. Prosodic stimuli were semantically neutral short sentences ( $M = 1511$  ms,  $SD = 305$ ; e.g., “Ele chega amanhã,” *He arrives tomorrow*) produced by two female and two male speakers (three different sentences per speaker per emotion). Peak amplitude was normalized to 75 dB using Praat (Boersma & Weenink, 2025). Facial stimuli were color photographs of male and female adults without beards, moustaches, earrings, eyeglasses, or visible makeup.

The procedure was identical for both tasks. After each stimulus was presented, participants selected the label that matched the expressed emotion from seven response options. Both tasks began with four practice trials. Prosodic stimuli were presented once; facial stimuli remained visible for 2 s. No feedback was provided. The order of the stimuli and response options were randomized across participants. Each trial consisted of a fixation cross (500 ms), followed by the stimulus and then the response. The two tasks took approximately 12 min to complete.

### Psychoacoustic tasks

Following Baldé et al. (2025), we administered nine psychoacoustic tasks from the PSYCHOACOUSTICS toolbox (Soranzo & Grassi, 2014). Three measured pitch perception using pure tones (pitch), tone sequences (contour speed), or musical scales (scale mistuning). Four measured temporal processing, including duration discrimination with pure tones (duration) and rhythmic sequences (complex duration), gap detection with broadband noise (gap detection), and temporal resolution with pure tones and noise (backward masking). The remaining two tasks measured loudness (intensity) and timbre perception (timbre). Sounds were presented at 75 dB SPL through headphones (Mars Gaming MH4X) and included 10-ms onset and offset ramps.

All but one task (scale mistuning) used a three-alternative forced-choice format. On each trial, three stimuli were presented successively: two identical standards and a comparison that differed on a single acoustic dimension. The comparison could be presented in the first, second, or third position. Listeners' task was to indicate the odd stimulus (1, 2, or 3), with target position randomized across trials. The difference between standards and comparison was adjusted adaptively using a two-down, one-up staircase procedure, targeting the minimum change required for 70.7% accuracy (Levitt, 1971; Soranzo & Grassi, 2014). Changes were salient at first to ensure task comprehension. After two consecutive correct responses, the difference between standards and comparison decreased by a factor of 2. After one incorrect response, it increased by a factor of 2. After four reversals, step size was reduced to  $\sqrt{2}$  to fine tune the estimate. Tasks ended after 12 reversals, with the threshold calculated as the geometric mean of the difference between standard and comparison stimuli on the final eight reversals. Lower thresholds indicated better performance, such that we expected negative associations between thresholds and other measures.

On each trial of the Pitch task, listeners heard three 250-ms pure tones and identified the one with the highest pitch. Standards were fixed at 1000 Hz and the comparison was higher. For contour speed, listeners heard three sequences of four tones and identified the one that had a contour change, which reversed the order of the second and third tones

(550–710 Hz or 710–550 Hz). The duration of the middle two tones decreased adaptively with correct responding, such that the test measured the shortest duration that still allowed for correct identification of a salient difference in pitch, as in tests of auditory inspection time (e.g., Deary, 1995). Scale mistuning had a yes/no response format that differed from the other tasks. Each trial presented a single ascending equal-tempered C-major scale (500-ms complexes of five harmonics). Listeners judged whether the scale was in or out of tune. Sol (G, the 5th scale degree) was mistuned upward to varying degrees.

Trials in the duration task comprised three pure tones. The two standards were 250 ms and the comparison was longer. For complex duration, trials contained three sequences of six 20-ms, 1000-Hz tone pulses arranged as three pairs. In the standards, silences were 40 ms within pairs and 120 ms between pairs. The comparison was identical in total duration, but within-pair silences were longer and between-pair silences were shorter. The magnitude of these differences decreased adaptively. For gap detection, listeners heard three 750-ms bands of Gaussian noise and identified the one containing a silent gap, which became briefer over time. For backward masking, each trial contained three 300-ms bands of bandpass noise and listeners identified which one had a 20-ms target tone that immediately preceded the noise. The target became quieter with correct responding.

Each trial in the Intensity task had three 1000 Hz pure tones (250 ms) and listeners identified the loudest. Finally, for Timbre, listeners heard three complex tones comprising five harmonics (e.g. 200, 400, 800, 1600, and 3200 Hz). Harmonics had equal amplitudes in standards, but the third harmonic was louder in the comparison.

### Self-reported music training and abilities

We assessed music training using the Music Training subscale from the Gold-MSI, a 38-item questionnaire that indexes different facets of musical expertise and behavior (Müllensiefen et al., 2014; Portuguese translation: Lima et al., 2020). Gold-MSI items are rated on 7-point scales and grouped into five subscales: Active Engagement (nine items, e.g., *I'm intrigued by musical styles I'm not familiar with and want to find out more*); Perceptual Abilities (nine items, e.g., *I can tell when people sing or play out of tune*), Music Training (seven items, e.g., *I had [0–10 or more] years of formal training on a musical instrument [including voice] during my lifetime*), Singing Abilities (seven items, e.g., *I only need to hear a new tune once and I can usually sing it back hours later*), and Emotions (six items, e.g., *Music can evoke my memories of past people and places*). In addition to an item asking about years of formal instruction, the Music Training subscale has additional items that ask about practice duration and frequency, compliments on musical

performances, and self-identification as a musician. We opted to use subscale scores in the analyses after confirming that they correlated highly with training duration (from the subscale's own training item; 0 = *no training*, 7 = *10 years or more*),  $r = .863, p < .001$ , and with responses to an additional question that asked about exact years of training (square-root transformed),  $r = .889, p < .001$ .

We also standardized and averaged the two subscales from the Gold-MSI that focused on musical abilities (Perceptual Abilities and Singing Abilities). The composite score is hereafter referred to as *Self-Reported Musicality*. The remaining two subscales (Active Engagement and Emotions) were irrelevant to our hypotheses and not considered further.<sup>1</sup>

### Objective musical abilities

We measured melody and rhythm perception objectively with the online version (Correia et al., 2022a, 2022b) of the MET (Wallentin et al., 2010). Because participants completed the test individually in the laboratory, testing closely matched the original computerized version. The test's first and second halves provided scores for the Melody and Rhythm subtests, respectively. Each subtest had 52 trials preceded by two practice trials. On each trial, listeners heard two short musical excerpts played on piano (Melody) or woodblock (Rhythm). Their task was to determine whether the first and second were identical (yes/no). In half of the trials, the second excerpt differed by changing the pitch of one or more tones in the Melody subtest, and one or more inter-onset intervals in the Rhythm subtest. No feedback was provided during test trials. Subtest scores were the number of correct responses. The entire MET took approximately 20 min to complete. Average performance accuracy ( $M = 69.1\%$ ) was similar to published norms ( $M = 69.7\%$ ; Swaminathan et al., 2021).

### Control measures

In addition to demographic variables (gender, age, and education; see Participants), we assessed general cognitive ability and personality.

**General cognitive abilities** Nonverbal (abstract) reasoning was measured with the Matrix Reasoning Item Bank (MaRs-IB; Chierchia et al., 2019), an online 8-min test modeled after Raven's Advanced Progressive Matrices (Raven, 1965).

<sup>1</sup> For completeness, associations between individual Gold-MSI subscales and the other primary measures (emotion recognition, psychoacoustic thresholds, objective musical abilities) are reported in the Supplementary Materials.

On each trial, participants had up to 30 s to respond, and scores were calculated as the proportion correct (number correct/trials attempted). Responses faster than 250 ms were excluded, and proportions were logit-transformed for statistical analysis.

Short-term and working memory were assessed with the forward and backward portions, respectively, of the Digit Span subtest from the Wechsler Adult Intelligence Scale–Third Edition (WAIS-III; Wechsler, 2008). In the forward portion, participants repeated aloud lists of numbers, of increasing length, in the presented order. In the backward portion, they recalled the digits in reverse order. Span length increased by one digit after every two trials. Testing stopped after failure on both trials of a given span. Scores on each portion were the maximum number of digits recalled correctly.

**Personality** The online version (Correia et al., 2022a, 2022b) of the Big Five Inventory (BFI; John et al., 2008; Portuguese translation: Brito Costa et al., 2016) evaluated the five major dimensions of personality (McCrae & John, 1992): Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness-to-Experience (hereafter, Openness). Participants rated 44 items on a 5-point scale (1 = *disagree strongly* to 5 = *agree strongly*). Each item represented a characteristic of one of the five dimensions (e.g., Openness: *I see myself as someone who values artistic, aesthetic experience*).

## Procedure

Participants were tested individually in a quiet room, with sessions lasting approximately 2 h. After obtaining informed consent, participants were interviewed briefly about their musical background. The tasks were then administered in the following order: first block of psychoacoustic tests (Pitch, Intensity, Duration, Complex Duration, Contour Speed), Digit Span, second block of psychoacoustic tests (Backward Masking, Gap Detection, Timbre, Scale Mistuning), Gold-MSI, BFI, MaRs-IB, MET, and finally the emotion-recognition tasks. The psychoacoustic tests were presented using MATLAB (r2022b, Version 9.13.0), with the order of tests randomized within each block for each participant. Subsequent questionnaires and tasks were implemented using the online platform Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Participants were allowed short breaks between tasks and received feedback about their musical abilities and personality at the end of the session.

## Data preparation and analysis

All analysis were conducted in JASP (Version 0.19.3; JASP Team, 2025). We report both standard frequentist (Null

Hypothesis Significance Testing [NHST]) and Bayesian statistics (default priors). For each analysis, we report a Bayes factor ( $BF_{10}$ ) with three-digit accuracy alongside  $p$  values. Whereas  $p$  values indicate the probability of the observed data under the null hypothesis, Bayes factors quantify the strength of evidence in favor of the alternative hypothesis.  $BF_{10}$  values  $> 1.00$  favor the alternative hypothesis. By convention (Jarosz & Wiley, 2014; Jeffreys, 1998),  $BF_{10} \leq 3.00$  indicates weak or anecdotal evidence;  $BF_{10} > 3.00$  indicates substantial evidence;  $BF_{10} > 10.0$ , strong evidence;  $BF_{10} > 30.0$ , very strong evidence; and  $BF_{10} > 100$ , decisive evidence. Conversely,  $BF_{10}$  values  $< 1.00$  favor the null hypothesis, with  $BF_{10} < .333$ ,  $.100$ ,  $.033$ ,  $.010$  corresponding, respectively, to substantial, strong, very strong, or decisive evidence for the null. We considered the results to provide reliable evidence for the alternative hypothesis only when  $p < .05$  and  $BF_{10} > 3.00$ , or for the null hypothesis when  $p > .05$  and  $BF_{10} < .333$ .

Because psychoacoustic thresholds typically follow a logarithmic distribution (Moore, 2003), they were log-transformed for statistical analyses (e.g., Baldé et al., 2025; Globerson et al., 2013; Micheyl et al., 2006). Thresholds and musical-ability scores deviating more than 3 standard deviations from the group mean (i.e., unusually high thresholds or unusually low musical-ability scores) were considered outliers and excluded. Missing data (MET Melody  $n = 1$ , Pitch  $n = 4$ , Intensity  $n = 2$ , Duration  $n = 3$ , Complex Duration  $n = 2$ , Contour Speed  $n = 9$ , Gap Detection  $n = 4$ , Timbre  $n = 8$ , Scale Mistuning  $n = 5$ ) meant that sample size varied across analyses depending on which variables were included.

Analyses of emotion recognition used average scores because we had no hypotheses about specific emotions, and because the literature typically reports a general advantage for musicians (e.g., Correia et al., 2022a, 2022b; Lima & Castro, 2011; Martins et al., 2021; Nussbaum & Schweinberger, 2021). Results for individual categories are nevertheless provided in Supplementary Materials.

## Results

Descriptive statistics are presented in Table 1. To identify relevant covariates, we examined correlations between demographic, cognitive, and personality factors and our primary measures of interest (emotion recognition, psychoacoustic thresholds, music training, musical abilities; Supplementary Tables S1–3). Age, gender, and education (demographics); nonverbal reasoning, digit span forward, and digit span backwards (cognition); and openness (personality) each correlated significantly with at least one primary measure. These seven factors were treated as *control variables* in all covariate-adjusted analyses.

**Table 1** Descriptive statistics

	<i>M</i>	<i>SD</i>
Demographics		
Age	22.7	7.28
Education (3 levels)	1.34	0.64
Emotion Recognition		
Prosody (%)	72.26	8.90
Faces (%)	83.16	5.92
MET		
Melody (%)	68.48	13.37
Rhythm (%)	69.79	11.01
Gold-MSI		
Music Training	3.30	1.82
Self-Reported Musicality	0.00	0.92
Psychoacoustic Thresholds		
Pitch (Hz)	23.91	32.18
Intensity (dB)	1.96	1.00
Duration (ms)	35.66	16.80
Complex Duration (ms)	53.24	54.88
Contour Speed (ms)	117.49	101.15
Backward Masking (dB)	-46.00	17.43
Gap Detection (ms)	2.33	0.64
Timbre (dB)	5.73	2.74
Scale Mistuning (semitones)	1.04	1.18
Cognitive abilities		
MaRs-IB–Nonverbal Reasoning (%)	64.70	14.28
Digit Span Forward	5.98	1.09
Digit Span Backwards	4.65	1.06
Personality		
Extraversion	3.17	0.82
Agreeableness	3.82	0.51
Conscientiousness	3.43	0.68
Neuroticism	3.26	0.78
Openness	3.78	0.63

The main analyses tested whether emotion recognition was associated with (1) music training, (2) musical abilities, and (3) basic auditory skills (psychoacoustic thresholds). We first examined each predictor separately using partial correlations adjusted for the control variables. We then implemented a joint model testing whether musical abilities remained predictive of prosodic emotion recognition when basic auditory skills were considered simultaneously. Secondary analyses compared highly trained ( $\geq 6$  years of lessons) and untrained ( $\leq 2$  years) participants to confirm that the correlational findings generalized to the analytical approach used commonly in the literature.

**Table 2** Partial pairwise correlations between music training (Music Training Subscale, Gold-MSI) and emotion recognition, musical ability, and psychoacoustic thresholds

Variable	<i>r</i>	<i>p</i>	BF <sub>10</sub>
Emotion Recognition			
Prosody	.031	.702	.289
Faces	-.030	.707	.288
Musical Ability			
Melody	<b>.520</b>	<b>&lt;.001</b>	<b>&gt;100</b>
Rhythm	<b>.364</b>	<b>&lt;.001</b>	<b>&gt;100</b>
Self-Reported Musicality	<b>.571</b>	<b>&lt;.001</b>	<b>&gt;100</b>
Psychoacoustic Thresholds			
Pitch	<b>-.452</b>	<b>&lt;.001</b>	<b>&gt;100</b>
Intensity	-.102	.207	.570
Duration	-.103	.206	.574
Complex Duration	-.041	.611	.306
Contour Speed	<b>-.312</b>	<b>&lt;.001</b>	<b>&gt;100</b>
Backward Masking	<b>-.313</b>	<b>&lt;.001</b>	<b>&gt;100</b>
Gap Detection	-.110	.175	.632
Timbre	-.072	.383	.394
Scale Mistuning	<b>-.310</b>	<b>&lt;.001</b>	<b>&gt;100</b>

Gender, age, education, nonverbal reasoning, digit span forward, digit span backwards, and openness were held constant. Significant findings are in bold font

## Music training

Table 2 reports partial correlations between music training (Gold-MSI subscale) and other primary measures after adjusting for control variables. As expected, music training correlated with better objective musical ability, higher self-reported musicality, and superior thresholds on pitch-based psychoacoustic tasks (Pitch, Contour Speed, Scale Mistuning) and Backward Masking.

Before covariate adjustment, music training correlated with prosodic emotion recognition,  $r = .194$ ,  $p = .013$ ,  $BF_{10} = 2.11$ , although Bayesian evidence was weak. For facial emotion recognition, there was substantial evidence for no association,  $r = .061$ ,  $p = .441$ ,  $BF_{10} = .131$ . After adjusting for control variables, the training-prosody association disappeared, and Bayesian evidence supported the null hypothesis (Table 2).

The same pattern emerged when highly trained ( $n = 50$ ) and untrained ( $n = 96$ ) groups were compared. For prosody, trained participants ( $M = 75.91\%$  correct,  $SD = 8.20$ ) outperformed untrained ones ( $M = 70.99$ ,  $SD = 8.53$ ) before covariate adjustment,  $t(144) = 2.29$ ,  $p = .024$ ,  $BF_{10} = 1.96$ , but Bayesian evidence was again weak. For faces, trained ( $M = 84.50$ ,  $SD = 6.31$ ) and untrained ( $M = 82.80$ ,  $SD = 5.62$ ) groups did not differ,  $t(144) = 1.35$ ,  $p = .180$ ,  $BF_{10} = .425$ . After covariate adjustment, an analysis of covariance (ANCOVA) confirmed that evidence for a null effect was

close to substantial for prosody,  $F < 1$ ,  $p = .593$ ,  $BF_{10} = .336$ , and substantial for faces,  $F < 1$ ,  $p = .471$ ,  $BF_{10} = .294$ . Thus, across correlational and group-based analyses, associations between music training and prosodic emotion recognition were weak and appeared to stem from individual differences in control variables.

## Musical abilities

Melody perception showed a substantial association with prosodic emotion recognition even after covariate adjustment,  $r = .211$ ,  $p = .008$ ,  $BF_{10} = 7.53$ , but not with facial emotion recognition,  $r = .076$ ,  $p = .345$ ,  $BF_{10} = .551$ . This melody-prosody association remained robust after controlling additionally for music training,  $r = .230$ ,  $p = .004$ ,  $BF_{10} = 14.0$  (Fig. 1a), confirming that it did not stem from greater training among higher performers. Rhythm perception, by contrast, did not correlate with prosody,  $r = .112$ ,  $p = .161$ ,  $BF_{10} = .816$ , or facial emotion recognition,  $r = .123$ ,  $p = .124$ ,  $BF_{10} = 1.04$ . Similarly, self-reported musicality showed no clear association with prosody,  $r = .141$ ,  $p = .077$ ,  $BF_{10} = 1.36$ , or faces,  $r = -.007$ ,  $p = .930$ ,  $BF_{10} = .375$ .

In short, objective musical ability, particularly melody perception, predicted prosodic emotion recognition independently of control variables and music training.

## Basic auditory skills

Table 3 presents partial correlations among auditory thresholds, musical abilities, and emotion recognition after accounting for control variables. Lower (better) thresholds across all nine psychoacoustic tasks predicted improved melody perception, with decisive evidence ( $BF_{10} > 100$ ) for pitch-related tasks (Pitch, Contour Speed, Scale Mistuning),

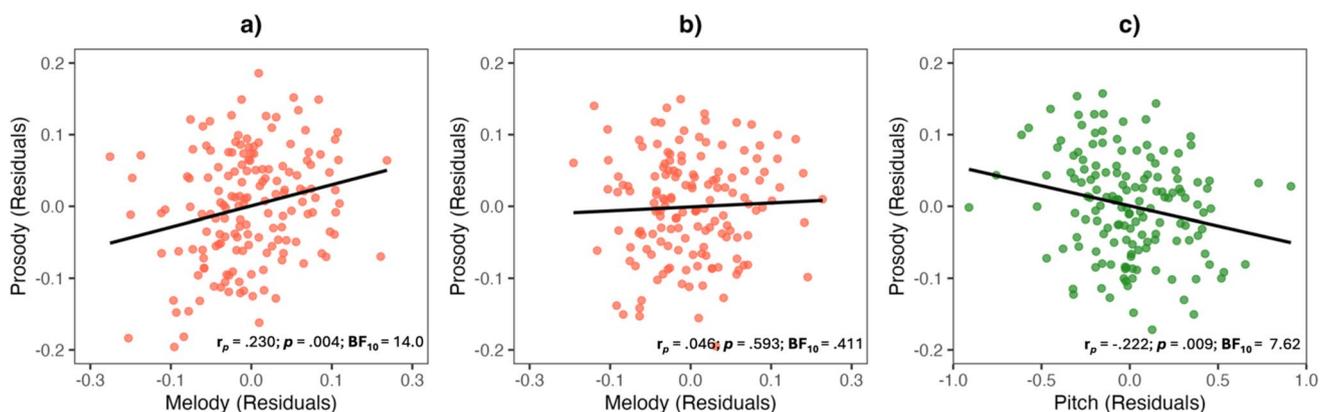
Backward Masking, and Timbre. Both rhythm and self-reported musicality were associated with seven psychoacoustic tasks. After adjusting additionally for music training, all associations between auditory thresholds and melody perception remained significant (Supplementary Table S5). Rhythm perception remained associated with only the Pitch and three temporal tasks (Duration, Complex Duration, and Backward Masking), whereas self-reported musicality was associated with six thresholds.

Four psychoacoustic thresholds predicted prosodic emotion recognition after covariate adjustment: Pitch (decisive evidence), Duration (very strong evidence), and Intensity and Gap Detection (strong evidence; Table 3). Correlations with facial emotion recognition were weaker, with substantial evidence only for Intensity and Duration. All associations remained significant after controlling additionally for music training (Supplementary Table S5).

Thus, auditory thresholds correlated with both musical abilities and prosodic emotion recognition, independently of control variables and music training.

## Joint prediction of emotional prosody from melody and basic auditory skills

Finally, we tested whether melody perception continued to predict prosodic emotion recognition when the psychoacoustic thresholds associated with prosody were considered simultaneously. A hierarchical multiple regression predicted prosody scores from control variables and music training at Step 1, adding at Step 2 melody and the four psychoacoustic thresholds (Pitch, Duration, Intensity, Gap Detection). Results are presented in Table 4. Step 1 explained 20.0% of the variance,  $R = .447$ ,  $F(8, 142) = 4.44$ ,  $p < .001$ . Significant predictors were higher openness, younger age,



**Fig. 1** Associations between prosodic emotion recognition and melody perception **a**) before and **b**) after adjusting for psychoacoustic thresholds (pitch, duration, intensity, gap detection). **c**) Association between prosodic emotion recognition and pitch perception after adjusting for melody perception and other psychoacoustic thresholds

(duration, intensity, gap detection). *Note.* In all panels, prosody, melody, and pitch were additionally adjusted for age, gender, education, music training, nonverbal reasoning, digit span forward, digit span backwards, and openness

**Table 3** Partial pairwise correlations between psychoacoustic thresholds, emotion recognition, and musical ability

		Prosody	Faces	Melody	Rhythm	Self-Reported Musicality
Pitch	<i>r</i>	−.339	−.158	−.568	−.327	−.439
	<i>p</i>	<.001	.051	<.001	<.001	<.001
	BF <sub>10</sub>	> 100	1.92	> 100	> 100	> 100
Intensity	<i>r</i>	−.229	−.205	−.253	−.193	−.281
	<i>p</i>	.004	.010	.002	.016	<.001
	BF <sub>10</sub>	13.2	6.40	29.4	4.17	88.8
Duration	<i>r</i>	−.279	−.178	−.245	−.275	−.131
	<i>p</i>	<.001	.028	.002	<.001	.105
	BF <sub>10</sub>	79.8	3.06	21.1	69.2	1.19
Complex Duration	<i>r</i>	−.158	−.067	−.274	−.217	−.124
	<i>p</i>	.049	.408	<.001	.007	.125
	BF <sub>10</sub>	1.90	.504	67.0	8.38	1.08
Contour Speed	<i>r</i>	−.073	−.086	−.469	−.276	−.274
	<i>p</i>	.381	.301	<.001	<.001	<.001
	BF <sub>10</sub>	.511	.607	> 100	59.2	52.8
Backward Masking	<i>r</i>	−.148	.101	−.407	−.317	−.339
	<i>p</i>	.065	.207	<.001	<.001	<.001
	BF <sub>10</sub>	1.55	.743	> 100	> 100	> 100
Gap Detection	<i>r</i>	−.240	.016	−.271	−.045	−.247
	<i>p</i>	.003	.844	<.001	.583	.002
	BF <sub>10</sub>	18.3	.384	54.6	.332	23.3
Timbre	<i>r</i>	−.149	−.068	−.315	−.145	−.237
	<i>p</i>	.070	.410	<.001	.078	.004
	BF <sub>10</sub>	1.51	.519	> 100	1.24	14.6
Scale Mistuning	<i>r</i>	−.130	.045	−.462	−.225	−.378
	<i>p</i>	.109	.582	<.001	.005	<.001
	BF <sub>10</sub>	1.08	.422	> 100	10.2	> 100

Gender, age, education, nonverbal reasoning, digit span forward, digit span backwards, and openness were held constant. Significant findings are in bold font

and female gender. On Step 2, explained variance increased by 14.3%,  $F(5, 137) = 5.97$ ,  $p < .001$ . Pitch discrimination emerged as the strongest independent predictor of prosodic emotion recognition. Melody perception no longer contributed unique variance, with Bayesian evidence slightly favoring the null hypothesis (Table 4; Fig. 1 b, c). Thus, when considered jointly, basic pitch sensitivity, but not melody perception, explained individual differences in prosodic emotion recognition.

## Discussion

Although links between musical expertise and speech processing are often attributed to music training (e.g., Patel, 2014; Slater et al., 2015), causal evidence is limited (Neves et al., 2022; Sala & Gobet, 2020; Schellenberg & Lima,

2024) and attempts to replicate frequently fail (Boebinger et al., 2015; Schellenberg et al., 2023; Whiteford et al., 2025). Recent work points instead to musical aptitude as the source of observed associations (Correia et al., 2022a, 2022b; Mankel & Bidelman, 2018; Swaminathan & Schellenberg, 2017). But what mechanisms link musical aptitude to speech processing? Our results answer this question by showing that (1) musical ability—particularly melody perception—predicts emotion recognition in speech prosody independently of music training, and (2) this association is explained by basic, domain-general pitch discrimination.

Bayesian analyses indicated that the apparent association between music training and prosodic emotion recognition was weak and disappeared after controlling for sociodemographic, cognitive, and personality covariates. This pattern suggests that associations reported in previous studies may reflect selection effects or confounding factors rather

**Table 4** Summary of hierarchical multiple regression model predicting prosodic emotion recognition

	$\beta$	$p$	$BF_{10}$	$r_p$	$R^2$	Adjusted $R^2$
<i>Step 1</i>					.200	.155
Age	−.236	.019	4.11	−.196		
Gender	.175	.024	3.37	.188		
Education	.045	.677	.384	.035		
Nonverbal Reasoning	.189	.041	2.27	.171		
Digit Span Forward	.103	.211	.712	.105		
Digit Span Backwards	.109	.228	.677	.101		
Openness	.204	.016	4.68	.201		
Music Training	.017	.854	.361	.016		
<i>Step 2</i>					.343	.281
Age	−.186	.058	1.78	−.161		
Gender	.197	.009	7.33	.221		
Education	.015	.882	.365	.013		
Nonverbal Reasoning	.106	.223	.700	.104		
Digit Span Forward	.045	.560	.420	.050		
Digit Span Backwards	.089	.290	.595	.090		
Openness	.192	.015	4.95	.206		
Music Training	−.183	.076	1.45	−.151		
Melody	.055	.593	.411	.046		
Pitch	−.280	.009	7.62	−.222		
Intensity	−.044	.616	.404	−.043		
Duration	−.116	.180	.803	−.114		
Gap Detection	−.139	.087	1.33	−.146		

Significant predictor variables are in bold font

than a causal role of training. Although some older studies reported training advantages that survived control for cognitive ability (Lima & Castro, 2011; Thompson et al., 2004), more recent research with children (Neves et al., 2025) and adults (Correia et al., 2022a, 2022b) failed to find effects after adjusting for relevant covariates. Training could, in principle, enhance prosody recognition indirectly via cognitive improvements, which would explain why associations disappear when cognitive abilities are controlled. In a recent randomized trial, however, 2 years of music lessons did not improve prosodic emotion recognition (Neves et al., 2025), and causal evidence that music training enhances cognitive abilities is weak or nonexistent (Bigand & Tillmann, 2022; Sala & Gobet, 2020; Schellenberg & Lima, 2024).

By contrast, objective musical ability, most clearly melody discrimination, remained a reliable predictor of prosodic emotion recognition after accounting for training and control variables. Crucially, this association was specific to the auditory domain. Melody perception predicted prosodic but not facial emotion recognition, pointing to an auditory-specific mechanism (Correia et al., 2022a, 2022b). Cross-modal effects may still occur in developing or atypical populations, however, a hypothesis that future research could examine further. In children, for example,

a correlation can emerge between musical ability and emotion recognition from faces, even though the link is stronger for prosody (Neves et al., 2025). Individuals with congenital amusia can also show deficits in both prosody and facial emotion processing (Lima et al., 2016; but see Zhishuai et al., 2017).

Our finding that melody, rather than rhythm, predicted prosodic emotion recognition replicates and extends prior evidence. In two studies, the melody subtest from the PROMS correlated with emotion recognition for pseudowords that conveyed emotions through multiple prosodic cues or pitch alone (Nussbaum et al., 2024, 2025). Reported associations with the rhythm subtest either disappeared when music training was controlled (Nussbaum et al., 2024) or were based on samples of trained participants (Nussbaum et al., 2025), which limits generalizability. In another study, children were tested with the Montreal Battery of Evaluation of Musical Abilities (Peretz et al., 2013) and the prosodic stimuli used here. Melody but not the rhythm subtest predicted prosodic emotion recognition (Neves et al., 2025). Altogether, these findings support the colloquial notion of prosody as the *melody of speech* and suggest that sensitivity to pitch changes over time is the principal contributor to the music–prosody overlap.

A key question concerns the processing level at which melodic and prosodic perception converge. The independence of the link from cognitive abilities, including short-term and working memory, argues against an explanation based on general memory for auditory sequences. Rather, our results point to low-level auditory sensitivity as the shared mechanism. Psychoacoustic thresholds correlated with both musical ability and prosodic emotion recognition. Consistent with Baldé et al. (2025) and models that place basic auditory processing at the foundation of musical abilities (Koelsch, 2011; Peretz & Coltheart, 2003), melody perception related strongly with pitch-related thresholds, whereas rhythm perception correlated with temporal and nontemporal thresholds, including pitch discrimination.

Although pitch discrimination emerged as the strongest predictor of prosodic emotion recognition, we also observed reliable associations with duration, intensity, and gap-detection thresholds, a finding in line with the notion that prosodic emotions are conveyed by multiple acoustic cues (Juslin & Laukka, 2003; Larrouy-Maestri et al., 2024). The prominent role of pitch is nevertheless consistent with results from Globerson et al. (2013), who found that two-tone pitch discrimination explained over 20% of the variance in prosodic emotion recognition. Similar findings have been observed in individuals with autism spectrum disorders (Globerson et al., 2015), and help to explain why emotion recognition remains high when prosody is conveyed solely by pitch cues (Nussbaum et al., 2024, 2025).

Importantly, when pitch thresholds and melody perception were considered simultaneously, only pitch discrimination contributed uniquely to explaining variance in a model that predicts prosodic emotion recognition. Higher-order melodic processing and other psychoacoustic predictors (duration, intensity, and gap-detection) failed to add explanatory power. Thus, even though multiple acoustic cues convey prosodic emotions, pitch discrimination appears to be the core building block shared by melody and prosody perception, which implies that the music–prosody link is primarily acoustic rather than music-specific. This interpretation is consistent with the auditory precision hypothesis (Saito, 2023; Saito & Tierney, 2025), which proposes that precise encoding of basic acoustic features underpins complex speech processing, including second-language learning.

But not all evidence aligns with this account. For example, Kempe et al. (2015) observed that musical ability predicted nonnative speech perception beyond pitch sensitivity (Kempe et al., 2015), and Nussbaum et al., (2024, 2025) reported that prosodic emotion recognition was more strongly correlated with melody than with a pitch test. Comparisons with Nussbaum et al.'s studies are difficult, however, because of methodological differences including smaller samples, a non-adaptive pitch task, and no direct comparison between melody and pitch. Future

research could examine whether different facets of speech perception are associated differentially with basic auditory thresholds and higher-order musical abilities, and test further whether results are contingent on methodological choices.

Neurophysiological methods such as EEG could also clarify which stages of pitch perception best explain recognizing emotions from prosody. Adaptive pitch-discrimination tasks reflect both subcortical auditory precision (Marmel et al., 2013) and the efficiency of cortical processing at early automatic and later attentional stages (Tervaniemi et al., 2005). Individuals with better pitch thresholds may recognize prosodic emotions more accurately because of more precise early encoding or, alternatively, because of more efficient use of pitch cues in attentional-evaluative processes. Some EEG studies suggest that musician differences in prosody perception emerge at later stages (> 500 ms; Lehnen et al., 2025; Nussbaum et al., 2023), whereas others implicate subcortical (Strait et al., 2009) or early cortical mechanisms (Pinheiro et al., 2015). To our knowledge, no EEG study has yet tested whether individual differences in pitch sensitivity predict prosodic emotion recognition.

As with any correlational design, causal direction remains unresolved. Although it is theoretically grounded to hypothesize that pitch sensitivity supports both musical and prosodic processing, extensive experience with speech could also enhance pitch perception over time, indirectly improving melody perception. In any event, our findings point to a clear hypothesis for future testing: basic pitch sensitivity explains links between musical expertise and prosodic emotion recognition. More generally, our results suggest that a low-level auditory mechanism is the primary locus of overlap between musical expertise and emotional prosody. Our findings also reinforce the view that music training rarely improves nonmusical abilities (Neves et al., 2025; Schellenberg & Lima, 2024).

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.3758/s13423-026-02865-z>.

**Funding** Open access funding provided by FCTIFCCN (b-on). This work was funded by the Portuguese Foundation for Science and Technology through a PhD studentship awarded to Aíssa Baldé (SFRH/BD/2022.12352) and a Scientific Employment Stimulus grant awarded to E. Glenn Schellenberg (CEECIND/03266/2018).

**Data availability** The dataset generated and analyzed during the current study is available in the Open Science Framework (OSF) repository ([https://osf.io/j692q/overview?view\\_only=fa4f6c681a444410b15b4f66752d49db](https://osf.io/j692q/overview?view_only=fa4f6c681a444410b15b4f66752d49db)).

**Code availability** Not applicable.

## Declarations

**Ethics approval** The study was approved by the ethics committee at ISCTE University Institute of Lisbon.

**Consent to participate** Participants provided written informed consent and received partial course credit or token remuneration.

**Consent for publication** Participants provided written informed consent.

**Conflicts of interest** The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Baldé, A. M., Lima, C. F., & Schellenberg, E. G. (2025). Associations between musical expertise and auditory processing. *Journal of Experimental Psychology: Human Perception and Performance*, 51(6), 747–763. <https://doi.org/10.1037/xhp0001312>
- Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 921–934. <https://doi.org/10.1037/a0021922>
- Bianchi, F., Santurette, S., Wendt, D., & Dau, T. (2016). Pitch discrimination in musicians and non-musicians: Effects of harmonic resolvability and processing effort. *Journal of the Association for Research in Otolaryngology*, 17(1), 69–79. <https://doi.org/10.1007/s10162-015-0548-2>
- Bigand, E., & Tillmann, B. (2022). Near and far transfer: Is music special? *Memory & Cognition*, 50(2), 339–347. <https://doi.org/10.3758/s13421-021-01226-6>
- Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America*, 137(1), 378–387. <https://doi.org/10.1121/1.4904537>
- Boersma, P., & Weenink, D. (2025). *Praat: Doing phonetics by computer* (Version 6.4.46) [Computer software]. <https://www.fon.hum.uva.nl/praat/>
- Brito Costa, S., Bem-Haja, P., Moisaio, A., Alberty, A., Vicente Castro, F., & De Almeida, H. (2016). Psychometric properties of Portuguese version of Big Five Inventory (BFI). *International Journal of Developmental and Educational Psychology*, 1(2), 83. <https://doi.org/10.17060/ijodaep.2015.n2.v1.325>
- Carton, J. S., Kessler, E. A., & Pape, C. L. (1999). Nonverbal decoding skills and relationship well-being in adults. *Journal of Nonverbal Behavior*, 23(1), 91–100. <https://doi.org/10.1023/A:1021339410262>
- Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods*, 42(1), 74–81. <https://doi.org/10.3758/BRM.42.1.74>
- Castro, S. L., & Lima, C. F. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception*, 32(2), 125–142. <https://doi.org/10.1525/mp.2014.32.2.125>
- Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S.-J. (2019). The matrix reasoning item bank (MaRS-IB): Novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science*, 6(10), Article 190232. <https://doi.org/10.1098/rsos.190232>
- Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2022a). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*, 22(5), 894–906. <https://doi.org/10.1037/emo0000770>
- Correia, A. I., Vincenzi, M., Vanzella, P., Pinheiro, A. P., Lima, C. F., & Schellenberg, E. G. (2022b). Can musical ability be tested online? *Behavior Research Methods*, 54(2), 955–969. <https://doi.org/10.3758/s13428-021-01641-2>
- Correia, A. I., Vincenzi, M., Vanzella, P., Pinheiro, A. P., Schellenberg, E. G., & Lima, C. F. (2023). Individual differences in musical ability among adults with no music training. *Quarterly Journal of Experimental Psychology*, 76(7), 1585–1598. <https://doi.org/10.1177/17470218221128557>
- Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, 4, Article 222. <https://doi.org/10.3389/fpsyg.2013.00222>
- Cowen, A., Sauter, D., Tracy, J. L., & Keltner, D. (2019). Mapping the passions: Toward a high-dimensional taxonomy of emotional experience and expression. *Psychological Science in the Public Interest*, 20(1), 69–90. <https://doi.org/10.1177/1529100619850176>
- Deary, I. J. (1995). Auditory inspection time and intelligence: What is the direction of causation? *Developmental Psychology*, 31(2), 237–250. <https://doi.org/10.1037/0012-1649.31.2.237>
- Endo, N., Ito, T., Watanabe, K., & Nakazawa, K. (2021). Enhancement of loudness discrimination acuity for self-generated sound is independent of musical experience. *PLoS ONE*, 16(12), Article e0260859. <https://doi.org/10.1371/journal.pone.0260859>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, 19(7), 573–576. <https://doi.org/10.1016/j.cub.2009.02.058>
- Globerson, E., Amir, N., Golan, O., Kishon-Rabin, L., & Lavidor, M. (2013). Psychoacoustic abilities as predictors of vocal emotion recognition. *Attention, Perception, & Psychophysics*, 75(8), 1799–1810. <https://doi.org/10.3758/s13414-013-0518-x>
- Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015). Prosody recognition in adults with high-functioning autism spectrum disorders: From psychoacoustics to cognition: Prosody recognition and psychoacoustics in ASD. *Autism Research*, 8(2), 153–163. <https://doi.org/10.1002/aur.1432>
- Grassi, M., Talamini, F., Altoè, G., Brattico, E., Caclin, A., Carretti, B., Draï-Zerbib, V., Ferreri, L., Gambarota, F., Grahn, J., Guiotto Nai Fovino, L., Rocco, M., Rodriguez-Fornells, A., Swaminathan, S., Tillmann, B., Vuust, P., Wilbiks, J., Zentner, M., Aguilar, K., ... Zappa, A. (2025). Do musicians have better short-term memory than nonmusicians? A multilab study. *Advances in Methods and Practices in Psychological Science*, 8(4), 1–30. <https://doi.org/10.1177/25152459251379432>

- 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>
- Jansen, N., Harding, E. E., Loerts, H., Başkent, D., & Lowie, W. (2023). The relation between musical abilities and speech prosody perception: A meta-analysis. *Journal of Phonetics*, *101*, Article 101278. <https://doi.org/10.1016/j.wocn.2023.101278>
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*(1). <https://doi.org/10.7771/1932-6246.1167>
- JASP Team. (2025). *JASP* (Version 0.19.3) [Computer software]. <https://jasp-stats.org/>
- Jeffreys, H. (1998). *Theory of probability* (3rd ed). Oxford University Press.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed, pp. 114–158). Guilford Press.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>
- Kempe, V., Bublitz, D., & Brooks, P. J. (2015). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology*, *106*(2), 349–366. <https://doi.org/10.1111/bjop.12092>
- Koelsch, S. (2011). Toward a neural basis of music perception—A review and updated model. *Frontiers in Psychology*, *2*, Article 110. <https://doi.org/10.3389/fpsyg.2011.00110>
- Larrouy-Maestri, P., Poeppel, D., & Pell, M. D. (2024). The sound of emotional prosody: Nearly 3 decades of research and future directions. *Perspectives on Psychological Science*, *20*(4), 623–638. <https://doi.org/10.1177/17456916231217722>
- Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS ONE*, *7*(12), Article e52508. <https://doi.org/10.1371/journal.pone.0052508>
- Lee, H., & Müllensiefen, D. (2020). The timbre perception test (TPT): A new interactive musical assessment tool to measure timbre perception ability. *Attention, Perception & Psychophysics*, *82*(7), 3658–3675. <https://doi.org/10.3758/s13414-020-02058-3>
- Lehnen, J. M., Schweinberger, S. R., & Nussbaum, C. (2025). Vocal emotion perception and musicality—insights from EEG decoding. *Sensors*, *25*(6), Article 1669. <https://doi.org/10.3390/s25061669>
- Leipold, S., Abrams, D. A., Karraker, S., & Menon, V. (2022). Neural decoding of emotional prosody in voice-sensitive auditory cortex predicts social communication abilities in children. *Cerebral Cortex*, *33*(3), 709–728. <https://doi.org/10.1093/cercor/bhac095>
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477. <https://doi.org/10.1121/1.1912375>
- Lima, C. F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S. K., Warren, J. D., & Stewart, L. (2016). Impaired socio-emotional processing in a developmental music disorder. *Scientific Reports*, *6*(1), Article 34911. <https://doi.org/10.1038/srep34911>
- Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021–1031. <https://doi.org/10.1037/a0024521>
- Lima, C. F., Correia, A. I., Müllensiefen, D., & Castro, S. L. (2020). Goldsmiths musical sophistication index (Gold-MSI): Portuguese version and associations with socio-demographic factors, personality and music preferences. *Psychology of Music*, *48*(3), 376–388. <https://doi.org/10.1177/0305735618801997>
- Lolli, S. L., Lewenstein, A. D., Basurto, J., Winnik, S., & Loui, P. (2015). Sound frequency affects speech emotion perception: Results from congenital amusia. *Frontiers in Psychology*, *6*, Article 1340. <https://doi.org/10.3389/fpsyg.2015.01340>
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). *Karolinska Directed Emotional Faces (KDEF)* [Database record]. <https://doi.org/10.1037/t27732-000>
- Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(51), 13129–13134. <https://doi.org/10.1073/pnas.1811793115>
- Marmel, F., Linley, D., Carlyon, R. P., Gockel, H. E., Hopkins, K., & Plack, C. J. (2013). Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. *Journal of the Association for Research in Otolaryngology*, *14*(5), 757–766. <https://doi.org/10.1007/s10162-013-0402-3>
- Martin, C. D., Niziolek, C. A., Duñabeitia, J. A., Perez, A., Hernandez, D., Carreiras, M., & Houde, J. F. (2018). Online adaptation to altered auditory feedback is predicted by auditory acuity and not by domain-general executive control resources. *Frontiers in Human Neuroscience*, *12*, Article 91. <https://doi.org/10.3389/fnhum.2018.00091>
- Martins, M., Pinheiro, A. P., & Lima, C. F. (2021). Does music training improve emotion recognition abilities? A critical review. *Emotion Review*, *13*(3), 199–210. <https://doi.org/10.1177/17540739211022035>
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*(1/2), 36–47. <https://doi.org/10.1016/j.heares.2006.05.004>
- Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). Academic Press.
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2), Article e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Neves, L., Correia, A. I., Castro, S. L., Martins, D., & Lima, C. F. (2022). Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. *Neuroscience and Biobehavioral Reviews*, *140*, Article 104777. <https://doi.org/10.1016/j.neubiorev.2022.104777>
- Neves, L., Martins, M., Correia, A. I., Castro, S. L., & Lima, C. F. (2021). Associations between vocal emotion recognition and socio-emotional adjustment in children. *Royal Society Open Science*, *8*(11), Article 211412. <https://doi.org/10.1098/rsos.211412>
- Neves, L., Martins, M., Correia, A. I., Castro, S. L., Schellenberg, E. G., & Lima, C. F. (2025). Does music training improve emotion recognition and cognitive abilities? Longitudinal and correlational evidence from children. *Cognition*, *259*, Article 106102. <https://doi.org/10.1016/j.cognition.2025.106102>
- Nussbaum, C., Dethloff, J., Schirmer, A., & Schweinberger, S. R. (2025). Vocal emotion recognition: A comparison of singers and instrumentalists, amateurs and professionals. *iScience*, *28*(12), Article 114089. <https://doi.org/10.1016/j.isci.2025.114089>
- Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2023). Electrophysiological correlates of vocal emotional processing in

- musicians and non-musicians. *Brain Sciences*, 13(11), Article 1563. <https://doi.org/10.3390/brainsci13111563>
- Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2024). Musicality-tuned to the melody of vocal emotions. *British Journal of Psychology*, 115(2), 206–225. <https://doi.org/10.1111/bjop.12684>
- Nussbaum, C., & Schweinberger, S. R. (2021). Links between musicality and vocal emotion perception. *Emotion Review*, 13(3), 211–224. <https://doi.org/10.1177/17540739211022803>
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, 2, Article 142. <https://doi.org/10.3389/fpsyg.2011.00142>
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, 308, 98–108. <https://doi.org/10.1016/j.heares.2013.08.011>
- Paz, Y., Sun, S., Flum, M., Rodriguez, Y., Brown, E., Plate, R. C., & Waller, R. (2025). Hearing a feeling: Music emotion recognition and callous-unemotional traits in early childhood. *Child Development*, 96(6), 2247–2256. <https://doi.org/10.1111/cdev.70024>
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6(7), 688–691. <https://doi.org/10.1038/nn1083>
- 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>
- Pinheiro, A. P., Vasconcelos, M., Dias, M., Arrais, N., & Gonçalves, Ó. F. (2015). The music of language: An ERP investigation of the effects of musical training on emotional prosody processing. *Brain and Language*, 140, 24–34. <https://doi.org/10.1016/j.bandl.2014.10.009>
- Pralus, A., Fornoni, L., Bouet, R., Gomot, M., Bhatara, A., Tillmann, B., & Caclin, A. (2019). Emotional prosody in congenital amusia: Impaired and spared processes. *Neuropsychologia*, 134, Article 107234. <https://doi.org/10.1016/j.neuropsychologia.2019.107234>
- Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception*, 24(1), 37–48. <https://doi.org/10.1525/mp.2006.24.1.37>
- Raven, J. C. (1965). *Advanced progressive matrices sets I and II*. Oxford Psychologists Press.
- Saito, K. (2023). How does having a good ear promote successful second language speech acquisition in adulthood? Introducing Auditory Precision Hypothesis-L2. *Language Teaching*, 56(4), 522–538. <https://doi.org/10.1017/S0261444822000453>
- Saito, K., & Tierney, A. (2025). Roles of domain-general auditory processing in second language speech learning revisited: What degree of precision makes a difference? *Language Learning*, 75(S1), 97–138. <https://doi.org/10.1111/lang.12722>
- Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, 48(8), 1429–1441. <https://doi.org/10.3758/s13421-020-01060-2>
- Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, 31(1), 97–113.
- Schellenberg, E. G. (2020). Correlation = causation? Music training, psychology, and neuroscience. *Psychology of Aesthetics, Creativity, and the Arts*, 14(4), 475–480. <https://doi.org/10.1037/aca000263>
- Schellenberg, E. G., Correia, A. I., & Lima, C. F. (2023). Is musical expertise associated with self-reported foreign-language ability? *Journal of Experimental Psychology: Human Perception and Performance*, 49(7), 1083–1089. <https://doi.org/10.1037/xhp0001116>
- Schellenberg, E. G., & Lima, C. F. (2024). Music training and nonmusical abilities. *Annual Review of Psychology*, 75, 87–128. <https://doi.org/10.1146/annurev-psych-032323-051354>
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76–92. <https://doi.org/10.1177/0022022101032001009>
- Singh, M., & Mehr, S. A. (2023). Universality, domain-specificity and development of psychological responses to music. *Nature Reviews Psychology*, 2, 333–346. <https://doi.org/10.1038/s44159-023-00182-z>
- Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioural Brain Research*, 291, 244–252. <https://doi.org/10.1016/j.bbr.2015.05.026>
- Soranzo, A., & Grassi, M. (2014). Psychoacoustics: A comprehensive MATLAB toolbox for auditory testing. *Frontiers in Psychology*, 5, Article 712. <https://doi.org/10.3389/fpsyg.2014.00712>
- Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency—Effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience*, 29(3), 661–668. <https://doi.org/10.1111/j.1460-9568.2009.06617.x>
- Swaminathan, S., Kragness, H. E., & Schellenberg, E. G. (2021). The musical ear test: Norms and correlates from a large sample of Canadian undergraduates. *Behavior Research Methods*, 53(5), 2007–2024. <https://doi.org/10.3758/s13428-020-01528-8>
- Swaminathan, S., & Schellenberg, E. G. (2017). Musical competence and phoneme perception in a foreign language. *Psychonomic Bulletin & Review*, 24(6), 1929–1934. <https://doi.org/10.3758/s13423-017-1244-5>
- Swaminathan, S., & Schellenberg, E. G. (2020). Musical ability, music training, and language ability in childhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(12), 2340–2348. <https://doi.org/10.1037/xlm0000798>
- Terracciano, A., Merritt, M., Zonderman, A. B., & Evans, M. K. (2003). Personality traits and sex differences in emotion recognition among African Americans and Caucasians. *Annals of the New York Academy of Sciences*, 1000(1), 309–312. <https://doi.org/10.1196/annals.1280.032>
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, 161(1), 1–10. <https://doi.org/10.1007/s00221-004-2044-5>
- Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 109(46), 19027–19032. <https://doi.org/10.1073/pnas.1210344109>
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, 4(1), 46–64. <https://doi.org/10.1037/1528-3542.4.1.46>
- Vidas, D., Dingle, G. A., & Nelson, N. L. (2018). Children's recognition of emotion in music and speech. *Music & Science*, 1, 1–10. <https://doi.org/10.1177/2059204318762650>
- Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, 20(3), 188–196. <https://doi.org/10.1016/j.lindif.2010.02.004>
- Wechsler, D. (2008). *WAIS-III: Manual da Escala de Inteligência de Wechsler para Adultos [WAIS-III: Wechsler Adult Intelligence Scale]* (3rd ed.). Cegoc-Tea [Hogrefe].

- Whiteford, K. L., Baltzell, L. S., Chiu, M., Cooper, J. K., Faucher, S., Goh, P. Y., Hagedorn, A., Irsik, V. C., Irvine, A., Lim, S.-J., Mesik, J., Mesquita, B., Oakes, B., Rajappa, N., Roverud, E., Schrlau, A. E., Van Hedger, S. C., Bharadwaj, H. M., Johnsrude, I. S., & Oxenham, A. J. (2025). Large-scale multi-site study shows no association between musical training and early auditory neural sound encoding. *Nature Communications*, *16*(1), Article 7152. <https://doi.org/10.1038/s41467-025-62155-5>
- Zhishuai, J., Hong, L., Daxing, W., Pin, Z., & Xuejing, L. (2017). Processing of emotional faces in congenital amusia: An emotional music priming event-related potential study. *NeuroImage: Clinical*, *14*, 602–609. <https://doi.org/10.1016/j.nicl.2017.02.024>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.