



OPEN

Prosodic discrimination skills mediate the association between musical aptitude and vocal emotion recognition ability

Julia Vigl[✉], Francesca Talamini, Hannah Strauss & Marcel Zentner

The current study tested the hypothesis that the association between musical ability and vocal emotion recognition skills is mediated by accuracy in prosody perception. Furthermore, it was investigated whether this association is primarily related to musical expertise, operationalized by long-term engagement in musical activities, or musical aptitude, operationalized by a test of musical perceptual ability. To this end, we conducted three studies: In Study 1 ($N=85$) and Study 2 ($N=93$), we developed and validated a new instrument for the assessment of prosodic discrimination ability. In Study 3 ($N=136$), we examined whether the association between musical ability and vocal emotion recognition was mediated by prosodic discrimination ability. We found evidence for a full mediation, though only in relation to musical aptitude and not in relation to musical expertise. Taken together, these findings suggest that individuals with high musical aptitude have superior prosody perception skills, which in turn contribute to their vocal emotion recognition skills. Importantly, our results suggest that these benefits are not unique to musicians, but extend to non-musicians with high musical aptitude.

Keywords Vocal emotion recognition, Musical aptitude, Musical expertise, Prosodic discrimination skills

Several studies have suggested a relationship between musical ability and the accurate recognition of vocal expressions of emotion (e.g.,^{1,2}). This association has been found primarily in the context of verbal stimuli³, occasionally extending to tone sequences that mimic the prosody of spoken emotional expressions⁴, and to nonverbal expressions of affect, such as infant distress⁵. Conversely, amusics show impaired recognition of vocal prosody, which has been attributed to difficulties in pitch perception⁶. The prevailing research trend suggests that musical ability is specifically associated with enhanced vocal emotion recognition rather than with emotion recognition in other domains (but see⁷). This suggests that the observed benefits of musical ability are primarily due to increased auditory sensitivity, rather than advantages in cognitive abilities, such as inferring emotion from emotional cues⁸. Consistent with this view, previous research has highlighted a link between musical ability and improved perception of prosodic signals in speech⁹.

However, although the relationship between musical ability and vocal emotion recognition has been extensively studied, the factors underlying this association are not fully understood. A factor that may play a mediating role is the ability to accurately perceive subtle changes in speech prosody, as prosody is crucial for conveying emotional nuances through variations in suprasegmental melodic and temporal properties of speech phonology.

Explaining the association between musical ability and emotion recognition skills

The domains of music and speech are closely intertwined, as evidenced by similarities in the way they are structured and how they are processed by listeners. These similarities may help explain the association found between musical ability and vocal emotion recognition (e.g.,^{1,2}). Specifically, one similarity lies in hierarchical organization, whereby the higher units are defined by melodies in music and phrases in speech, while subordinate units are defined by the musical sounds and speech phonemes, both of which are processed through the same auditory

Department of Psychology, University of Innsbruck, Universitätsstraße 15, 6020 Innsbruck, Austria. ✉email: julia.vigl@uibk.ac.at

pathways^{10–12}. Thus, the advantage in vocal emotion recognition found among musically proficient individuals may be explained by shared neural areas and operations for syntactic processing in music and language¹³.

Another possible explanation is that individuals with higher musical ability may have an enhanced perception of vocal modulations in speech, including those modulations that are associated with the expression of emotion¹. This explanation is supported by research showing that, in both music and in vocal expression, emotional information is expressed through variations in characteristics such as loudness, pitch and pitch contour, timbre, tempo, rhythm, or articulation^{14,15}. For example, in both domains, happiness is characterized by fast tempo and speech rate and a medium to high sound level and vocal intensity, whereas sadness is associated with slow tempo and speech rate and low sound level and intensity. Furthermore, there are parallels in pitch patterns used to express emotions, such as a descending minor third to express sadness or an ascending minor second to express anger^{16,17}, which could help explain why individuals who excel at recognizing emotions in speech have also been found to have an enhanced ability to recognize emotions in music¹⁸. The relevant modulations in speech are often subsumed under the term prosody, which refers to suprasegmental features of the voice accompanying speech acts. These include acoustic elements (loudness, pitch), stress (intonation curves, accentuation), speaking rate, speaking style (staccato, syllables separation), and aspects of verbal planning, such as hesitation, restatement, or stuttering¹⁹.

The idea that prosodic discrimination skills might mediate the association between musical ability and vocal recognition of emotion is consistent with a number of findings showing that speech-related perceptual skills can be enhanced by music training. Longitudinal studies in developmental psychology indicate that children engaged in musical training experience improvements in vocal intonation and pitch perception²⁰, syllable duration perception²¹, speech segmentation²², word discrimination²³, and phonological awareness^{24,25}. In addition, experimental studies have demonstrated that musicians have an heightened ability to detect pitch and timbre changes in both music and speech^{26,27} and advantages in speech-in-noise recognition²⁸. Overall, two recent meta-analyses showed a small positive effect of musical training on auditory processing²⁹, and a medium positive effect on linguistic and emotional speech prosody perception⁹. Given the importance of sensory processing in vocal emotion recognition^{8,30}, these findings are particularly relevant, as evidenced by the fact that individuals with hearing loss have difficulty associating auditory cues with emotions³¹.

However, it is at present unclear whether prosodic discrimination skills mediate the relationship between musical ability and the vocal recognition of emotion. Indeed, the studies included in the meta-analysis by Jansen et al.⁹ have not treated prosodic discrimination and vocal emotion recognition as distinct components within a mediation model. In our study, we define prosodic discrimination as the detection of subtle speech variations, whereas emotion recognition focuses specifically on the identification of vocally conveyed emotions.

Musical expertise versus aptitude

Early studies of the relationship between musical ability and vocal emotion recognition typically compared musicians and non-musicians, thus testing the effect of musical *expertise*, i.e., the effect of long-term engagement with music. However, dispositional skills in the perception of music, which are part of musical *aptitude*, do not necessarily require many years of formal musical training. Conversely, individuals with extensive musical training do not always demonstrate above-average music perception skills^{32–34}. In recent years, research has increasingly moved away from the practice of contrasting musicians and non-musicians toward investigating the effect of musical aptitude as a continuous variable. For example, musical aptitude was more strongly associated with speech-in-noise perception³⁵, vocal emotion recognition³⁶, and according to a recent meta-analysis, to prosody perception⁹, than musical expertise. For this reason, and to assess musical ability more comprehensively, both musical expertise and the performance in a musical aptitude battery are included in the present studies.

Components of musical aptitude relevant to emotional speech prosody

In striving to explain the association between musical ability and vocal emotion recognition in terms of prosodic discrimination skills, the question arises as to which components of music perception would be particularly relevant for prosodic discrimination and emotion recognition. Since not all music features have an obvious counterpart in speech prosody, we anticipate that the ability to identify emotion in vocal expressions will be associated with specific sub-abilities related to the central parameters of vocal emotion expression, rather than with general music perception skills. These parameters can be assigned to the categories pitch, loudness, temporal aspects, and voice quality^{10,37,38}.

More specific acoustic cues related to these categories, according to Juslin and Scherer³⁷, are components such as fundamental frequency, speech contour, and pitch jumps (which are components of pitch); speech intensity, rapidity of voice onset, and shimmer (related to loudness); speech rate, number of pauses, stressed syllables, and speech rhythm (related to temporal aspects); and high and low frequency energy in the spectrum, jitter, articulatory precision, and the slope of spectral energy (components of voice quality). Although these features do not operate in isolation, and their relevance varies from emotion to emotion, overall, the evidence described points to melody, tempo, rhythmic accents, and timbre perception as musical dimensions that should play a particularly prominent role in prosodic discrimination skills, and in turn also in the accuracy of vocal emotion recognition.

Measurement of prosodic discrimination skills

In order to explore the potential mediating role of prosodic discrimination in the association between musical ability and vocal emotion recognition, both components need to be measured objectively and reliably. Whereas some psychometrically sound vocal emotion recognition tests exist (e.g.,^{39,40}), we were unable to locate any instruments for the assessment of individual differences in prosodic discrimination ability when we started our research. Although some studies have used tasks to assess whether participants can discriminate between

intonations used in statements as opposed to questions (linguistic prosody, e.g.,^{41,42}), or infer emotions directly from speech samples (emotional prosody, e.g.,⁴³), there appears to be a lack of psychometrically sound, construct-validated test instruments that assess prosodic discrimination skills at a lower level, namely the ability to detect subtle prosodic differences in spoken sentences. Thus, in developing the present tool for assessing accuracy in perception of subtle changes in speech prosody in Studies 1 and 2, we prioritized discriminant validity as part of construct validation, so as to ensure that the tool measures prosodic discrimination skills rather than general auditory discrimination skills. The new instrument requires participants to detect variances both in parametric manipulations, such as pitch variations, and emotional manipulations, by changing the emotional coloration through multi-parametric adjustments.

The current studies

In the current research, we hypothesized that the association between musical ability and vocal emotion recognition would be mediated by prosodic discrimination skills. We expected that the postulated relationships would relate specifically to musical aptitude (rather than expertise), and even more specifically to the perception of those musico-acoustic parameters that are responsible for the modulation of prosodic features, namely pitch and pitch contour (i.e., melody), timbre, tempo, and rhythmical accents^{10,37,38}.

In Studies 1 and 2, we created and validated a pool of stimuli designed to measure the ability to detect subtle prosodic changes in speech recordings, as the basis for a new test to assess prosodic discrimination skills. In Study 3, we tested the mediation hypothesis, and whether musical aptitude makes a stronger contribution to vocal emotion recognition than expertise. All studies were approved by the Ethics Committee of the University of Innsbruck (Certificate of good standing, 25/2022), the methods were performed in accordance with the relevant guidelines and regulations, and each participant gave informed consent prior to participation consistent with the Declaration of Helsinki.

Study 1

The purpose of this study was to develop and validate stimuli for assessing prosodic discrimination ability. It focused on item creation, item analysis, and item reduction.

Method

Participants

Eighty-five German-speaking participants (75.3% female, mean age = 24.82, $SD = 9.39$) tested the initial pool of 48 stimuli (see *Method* section). Most of participants had finished general secondary education (42.4%) or had a university degree (36.5%).

Stimuli to assess prosodic discrimination skills

To measure prosodic discrimination ability, we generated a task in which participants are required to determine if a test stimulus sounds same or different compared to the reference stimulus. For our initial item pool, we used 48 recordings of neutral content sentences provided by Arias et al.⁴⁴, spoken in both neutral and emotional prosody, multiple languages and by male and female speakers. We included both neutral and emotional prosody to increase the ecological validity of the test, while also refraining from an excessive range of emotions to avoid making the test too similar to a vocal emotion recognition test.

We used 16 of these 48 items as "same" trials, meaning that they were included unchanged in our new test instrument. Half of them were recorded in neutral prosody, and the other half in emotional prosody, such as sad or happy, or fearful expressions.

To create the 32 "different" trials, we used the DAVID software⁴⁵, which allows for precise modification of vocal signals. We modified the audio recordings in two ways. First, for 16 items, we applied parametric changes to individual or combined characteristics, such as pitch, inflection, and vibrato, without introducing emotional coloration. We achieved this by employing alternative recordings and modifying pitch and inflection. Second, for the remaining 16 items, we altered the emotional prosody by applying emotional speech transformation templates provided by the software, incorporating sound effects associated with specific emotional qualities. Thereby we changed neutral sentences to express emotions (emotional coloring), intensified already emotionally spoken sentences in their expressed emotion (emotional intensification), or changed them in their expressed emotion (emotional switch).

All audio files were exported as MP3 files and normalized to a constant volume. In total, we obtained a set of 48 stimuli, including 16 unmodified and 32 modified stimuli (see Table 1A, for an overview). Each trial consisted of three audio recordings created with Audacity⁴⁶: First the reference stimulus, followed by its repetition 1.5 s later, and then the comparison stimulus after 2.5 s, which was either the same or different. The reference stimulus was presented twice to facilitate its encoding, thereby leaving less room for individual differences in memory capacity to affect the performance (for more detail on stimuli see Supplementary Materials, Section 1).

Procedure

We used LimeSurvey⁴⁷ to deliver the 48 stimuli online in a randomized order. We chose to offer five response options for each stimulus, going beyond the simple "same" versus "different" distinction to allow for more fine-grained sensory judgments⁴⁸. First, in line with signal detection theory⁴⁹, we included the participants' confidence level with the options "definitely same/definitely different" and "probably same/probably different". Second, we included the additional option "I don't know" to avoid guessing. Participants received 1 point for each correct high confidence answer (i.e., "definitely same/different") and 0.5 points for each correct low confidence answer (i.e., "probably same/different"). Incorrect and "I don't know" answers were given 0 points. The total score was

A: first version before item reduction		B: second version after item reduction	
16 “same” trials	8 neutrally spoken sentences 8 emotionally spoken sentences	8 “same” trials	4 neutrally spoken sentences 4 emotionally spoken sentences
32 “different” trials	16 parametric changes: 4 alternative recordings 4 pitch modifications 4 inflection modifications 4 pitch and inflection modifications	18 “different” trials	9 parametric changes: 2 alternative recordings 2 pitch modifications 2 inflection modifications 3 pitch and inflection modifications
	16 emotional prosody changes: 4 items with emotional switch 6 emotionally colored items 6 emotionally intensified items		9 emotional prosody changes: 3 items with emotional switch 3 emotionally colored items 3 emotionally intensified items
	Total: 48 trials	Total: 26 trials	

Table 1. Overview of the different items in the first and modified version of the prosody test.

calculated by summing all responses. In addition, we obtained informed consent, included demographic questions and provided a comment box for participants to give feedback on the prosodic discrimination task.

Results and discussion

Applying general principles of item analysis⁵⁰, we retained 26 trials that were balanced in terms of languages used, gender of speakers, parametric or emotional modifications, and difficulty, as shown in Table 1B (see Supplementary Materials, Section 2 and Table S2, for details). Of the modified stimuli, nine were edited with respect to individual or combined parameters and nine were edited with emotional speech transformations provided by the DAVID software. After item reduction, emotional prosody was restricted to happy and sad expressions.

Internal consistency of the test was $\omega = 0.84$, the average total score was 17.78 ($SD = 4.27$, range = 7.5–26), and item difficulty ranged from 0.34 to 0.90 ($M = 0.68$; $SD = 0.19$). Since the test included both purely parametric and emotional components of prosody that could potentially have different mediating roles, we also examined the internal consistencies of the two subtests, which were $\omega = 0.71$ and $\omega = 0.72$, respectively. We created subtest scores relating to the two components and found them to be significantly intercorrelated ($r = 0.62$, $p < 0.001$). There were no significant differences in item mean scores between the four languages or between stimuli with male or female speakers, and no significant differences in total scores based on individual characteristics such as gender, education, age, or the use of headphones or loudspeakers during participation (all $ps > 0.05$).

Sensitivity was estimated using Vokey’s⁵¹ d_p , which is obtained by fitting receiver operating characteristic (ROC) curves through principal component analysis and provides a robust estimate for confidence ratings compared to traditional sensitivity measures such as d' or d_a ^{51,52}. Average d_p was 1.63 ($SD = 0.67$). The total raw scores and d_p values were normally distributed, with $D(85) = 0.09$, $p = 0.562$ and $D(85) = 0.05$, $p = 0.958$, respectively. We found a strong correlation between the total raw score and the d_p value ($r = 0.94$, $p < 0.001$).

In summary, Study 1 led to a novel 26-item instrument to measure prosodic discrimination ability with two subtests, that demonstrated good initial psychometric characteristics overall.

Study 2

In this study, we examined the convergent and discriminant validity, and retest reliability of the prosodic discrimination test created in Study 1.

To establish convergent validity, we measured music perception skills using the Micro-PROMS⁵³, expecting moderate correlations due to the associations between musical aptitude and auditory and linguistic perception²⁹. For discriminant validity, we used two auditory tests that assess sensory thresholds for frequency discrimination and silent gap detection⁵⁴. In accordance with general criteria for demonstrating discriminant validity⁵⁵, we expected some association between prosodic discrimination skills and auditory thresholds, strongest for frequency discrimination, as the task primarily involved pitch modulation. Conversely, we expected a weaker correlation with the gap-in-noise task, as it does not rely on pitch discrimination.

Method

Participants

A total of 93 individuals participated in the first part of the study using headphones (62% female, mean age = 23.45, $SD = 3.53$). All of these individuals completed the prosodic discrimination test and the Micro-PROMS at the initial assessment, 62 further completed both auditory tests, and 64 participated in the retest. The difference in participant numbers between the auditory tests and the PROMS can be attributed to the requirement for participants to navigate away from the primary survey via two external links for the auditory tests.

Measures and procedure

The study was conducted online using LimeSurvey⁴⁷ and took approximately 30 min to complete. Two weeks after the first assessment, participants were invited to complete the prosodic discrimination task again for test–retest reliability.

For convergent validity, a novel and short version of the *Profile of Music Perception Skills*, the Micro-PROMS⁵³, was employed. This test contains a total of 18 items covering all the subtests of the full version of PROMS³⁴ and

can be completed in 10 min. Participants were asked to indicate whether a test stimulus sounds the same or different from a reference stimulus presented twice. The items had sufficient internal consistency ($\omega = 0.64$).

Two auditory tests were administered to assess discriminant validity: the *White Noise Gap Detection Test* and the *Pure Tone Frequency Discrimination Test*. These tests are classic psychoacoustic tasks that measure the sensory thresholds for frequency discrimination and silent gap detection using an adaptive procedure (e.g.,⁵⁶) and are available on a newly developed website (<http://psychoacoustics.dpg.psy.unipd.it/sito/index.php>) based on the MATLAB psychoacoustic toolbox⁵⁴. In both tests, a low threshold indicates better auditory performance. A detailed description of both tasks can be found in the Supplementary Material, Section 3.

Results and discussion

For the prosodic discrimination test, the total score distributions, d_p values and internal consistency were similar to those observed in Study 1 (see Table S2). In addition, we obtained high test–retest reliability ($r = 0.88$, $p < 0.001$, ICC = 0.92, $n = 64$).

In terms of convergent validity, we found a moderately strong correlation between prosodic discrimination and the Micro-PROMS ($r = 0.62$, $p < 0.001$, $n = 93$). For discriminant validity, there was a moderate correlation between prosodic discrimination and the *Pure Tone Frequency Discrimination Test* ($r = -0.33$, $p = 0.010$, $n = 62$) and a non-significant weak correlation between prosodic discrimination and the *White Noise Gap Detection Test* ($r = -0.10$, $p = 0.462$, $n = 62$). These results suggest that, although there is some overlap between auditory perception and prosodic discrimination, they are distinct constructs, and our measure predominantly taps into the latter. Reliability, test–retest statistics and criterion correlations for the two subtests of the prosodic discrimination test were largely comparable (see Table S2).

Study 3

In Study 3, we hypothesized that (1) both musical aptitude and expertise would be positively associated with vocal emotion recognition, with aptitude having a stronger association than expertise (e.g.,^{35,36}); (2) musical aptitude would be positively associated with prosodic discrimination skills; (3) prosodic discrimination skills would be positively associated with vocal emotion recognition; and (4) the association between musical aptitude and vocal emotion recognition would be mediated by prosodic discrimination skills.

To particularize the advantage of musical ability in vocal emotion recognition (e.g.,³⁶; but see also⁷), we use a multimodal emotion recognition test with auditory, visual, and audiovisual stimuli³⁹. Musical aptitude was assessed using a multicomponent battery³⁴; see “Method” section), which allowed identification of the components most strongly associated with prosodic discrimination and vocal emotion recognition ability. We expected the strongest correlations with the melody, pitch, timbre, accent, and tempo subtests, based on previous literature^{10,37,38}.

Method

Participants

A total of 136 participants without hearing impairment took part in the study (61.8% female), with a mean age of 24.28 years ($SD = 7.79$, range = 18–66). The majority of participants had completed high school (74%) or university (15%). Sixty-eight percent of the participants lived in Austria, 16% in Germany, and 17% in Italy. About half of the participants (56.6%) identified themselves as non-musicians, while 43.4% identified themselves as musicians (37 amateur musicians, 21 (semi)-professional musicians). Just over half of the participants play at least one instrument or sing (57.4%), for a mean of 11.45 years ($SD = 8.83$, range = 1–62). Amateur musicians ($n = 77$) reported practicing 3.97 h ($SD = 6.50$) per week, while (semi)-professional musicians ($n = 22$) reported practicing 9.34 h ($SD = 6.85$). Self-reported non-musicians and musicians (i.e., amateur, semi-professional or professional) did not differ in age and educational level (all $ps > 0.05$).

Measures

Prosodic discrimination skills. Internal consistency of the test developed in Studies 1 and 2 was satisfactory ($\omega_{Total} = 0.82$; $\omega_{Parametric} = 0.67$; $\omega_{Emotional} = 0.69$), mean item difficulty was appropriate (0.63), and the total scores ($M = 16.32$, $SD = 4.25$) were normally distributed according to a Kolmogorov–Smirnov test, $D(136) = 0.06$, $p = 0.799$. Sensitivity was again estimated using Vokey’s⁵¹ d_p , yielding an average value of 1.38 ($SD = 0.66$).

Emotion recognition ability. The *Emotion Recognition Assessment in Multiple Modalities Test* (ERAM³⁹) was used to assess participants’ ability to recognize emotions in audio, visual, and audiovisual presentations. The test consisted of 72 items from the GEMEP corpus, a collection of video clips of emotional expressions in pseudo-linguistic sentences⁵⁷. In each of the three subtests (audio, visual, audiovisual), participants had to indicate which of 12 (pre)selected emotions was presented. The emotions included hot anger, anxiety, despair, disgust, panic fear, happiness, interest, irritation, pleasure, pride, relief, and sadness. In the audiovisual condition, full videos were presented, while in the audio subtest only the audio track of the videos was played and in the visual subtest the videos were shown without sound. The internal consistency of the subtests, calculated using the Kuder–Richardson Formula 20 for dichotomous data⁵⁸, was lower than in the original study ($\alpha_{total} = 0.80$ and no internal consistencies reported for the subtests³⁹), with $\omega_{total\ score} = 0.69$, $\omega_{audio} = 0.50$, $\omega_{visual} = 0.28$, and $\omega_{audiovisual} = 0.50$ in the present study. Given the particular importance of vocal emotion recognition in this study, we examined the low reliability of the audio subtest and identified four items with very low (< 0.05) or negative item–total correlations. These four items were excluded prior to score calculation ($\omega = 0.54$).

Musical aptitude. Musical aptitude was assessed using a short version of the *Profile of Music Perception Skills* (PROMS-S⁵⁹). As in the full version, different aspects of music perception are tested with the eight subtests melody, tuning, tempo, accent, rhythm, embedded rhythm, pitch, and timbre. Participants listen to a reference stimulus twice and then decide whether a target stimulus is the same or different, with the same answer format and scoring as in the Micro-PROMS⁵³. The internal consistency of the total score was $\omega = 0.87$, whereas subtest scores ranged from $\omega = 0.44$ (timbre) to $\omega = 0.64$ (embedded rhythm).

Musical expertise. Musical expertise, as a person's musical background and training, was assessed through five music-specific questions. Participants were asked about their musical self-assessment (1 = *non-musician*, 2 = *music-loving non-musician*, 3 = *amateur musician*, 4 = *semi-professional musician*, 5 = *professional musician*), whether they played an instrument or sang and, if so, for how many years, how many hours per week they practiced, and whether they had graduated from a music university or conservatory. As these questions had a high internal consistency ($\omega = 0.90$), they were z-transformed and combined into one measure of musical expertise.

Procedure

The study was conducted online using the LimeSurvey software⁴⁷. Participants were recruited through the university mailing list, flyers and posts on social networks of music universities and conservatories. Psychology students received course credits for their participation, and musicians were compensated with €10. After answering demographic and music-specific questions at the beginning of the study, participants completed the prosodic discrimination test and were then referred to the ERAM and PROMS-S. In total, the study took approximately 75 min to complete.

Data analysis and power

We calculated a mediation model using the PROCESS macro in SPSS (version 4.0⁶⁰). Preacher and Hayes' bias-corrected nonparametric bootstrapping technique with 5000 bootstrap samples were used to estimate direct and indirect effects⁶¹. The web-based *Monte Carlo Power Analysis for Indirect Effects* application (https://schoemanna.shinyapps.io/mc_power_med/) was consulted to determine the required sample size for mediation assumptions⁶². Small to medium effects were expected for each mediation pathway. A minimum sample size of 133 subjects was required to achieve a power of 0.80.

Results

Descriptive results

Table 2 presents descriptive statistics and correlations between the variables relevant to our hypotheses. Participants were most successful at emotion recognition when the presentation was audiovisual, followed by visual and auditory presentations. Participants using headphones to participate ($n = 96$) did not differ from those using speakers ($n = 40$) on any of the test instruments used (all $ps > 0.05$). The pattern of correlations between musical aptitude, musical expertise, prosodic discrimination, and vocal emotion recognition conformed to expectations. In particular, only musical aptitude was significantly associated with vocal emotion recognition ability, whereas expertise was not. Regarding the two subtests of the prosodic discrimination test, the correlations with musical expertise, musical aptitude, and vocal emotion recognition did not differ significantly from each other, as determined by z-tests (see Table S3 for details).

To particularize the unique contribution of aptitude as opposed to expertise, we conducted a subgroup analysis contrasting individuals with high and low aptitude within the low expertise group, as well as those with high and low expertise within the low aptitude group. The groups were derived by median splits. T-tests showed that among individuals with low musical expertise ($n = 68$), those with high musical aptitude ($M = 66.25$, $SD = 11.57$) were significantly better at vocal emotion recognition than those with low aptitude ($M = 57.60$, $SD = 16.28$), $t(66) = -2.16$, $p = 0.035$, $d = -0.57$. Conversely, among those with low musical aptitude ($n = 71$), there was no significant difference between participants with low ($M = 57.60$, $SD = 16.28$) and high ($M = 59.57$, $SD = 13.14$) musical training, $t(69) = -0.50$, $p = 0.616$, $d = -0.13$. This analysis highlights that so-called "musical sleepers"³⁴, i.e., untrained individuals with high musical aptitude, can also show advantages in emotion recognition, unlike trained individuals with low aptitude.

As shown in Table 3, we explored which specific components of musical aptitude were particularly associated with prosodic discrimination and emotion recognition. We found strong correlations between several PROMS-S

Outcome	M (SD)	1	2	3	4	5	6
1. Musical expertise	0.00 (0.82)	[0.90]					
2. Musical aptitude	44.91 (8.37)	0.47**	[0.87]				
3. Prosodic discrimination	16.32 (4.25)	0.28**	0.62**	[0.82]			
4. Visual emotion recognition	62.48 (9.85)	0.20*	0.08	0.15*	[0.28]		
5. Audiovisual emotion recognition	71.60 (11.38)	0.16*	0.17*	0.19*	0.28**	[0.50]	
6. Vocal emotion recognition	60.96 (14.69)	0.13	0.22*	0.27**	0.43**	0.43**	[0.54]

Table 2. Descriptive statistics, zero-order correlations, and reliability statistics (McDonald's Omega) presented in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Variable	Prosodic discrimination	Vocal emotion recognition	Visual emotion recognition	Audiovisual emotion recognition
Musical expertise	0.28*	0.13	0.20*	0.16 [†]
PROMS total score	0.62**	0.22**	0.08	0.17 [†]
Melody	0.47**	0.23**	0.13	0.07
Rhythm	0.28**	0.14 [†]	0.03	- 0.01
Embedded rhythm	0.46**	0.16 [†]	0.08	0.21*
Tuning	0.41**	0.08	- 0.01	0.14
Accent	0.31**	0.06	0.02	0.12
Timbre	0.42**	0.20*	0.03	0.12
Tempo	0.54**	0.17 [†]	0.12	0.14
Pitch	0.44**	0.15 [†]	0.03	0.10

Table 3. Correlations between musical ability (expertise and aptitude) and prosodic discrimination ability and emotion recognition. [†] $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

subtests and prosodic discrimination, with the lowest correlation for accent and the highest for tempo. Vocal emotion recognition was found to be positively associated with the melody and timbre subtests, and marginally significant correlations were found for the rhythm, embedded rhythm, and tempo subtests. Finally, there were no correlations between visual emotion recognition and the PROMS-S subtests, while audiovisual emotion recognition was only correlated with the embedded rhythm subtest.

The first three hypotheses, namely that there are correlations between musical aptitude, vocal emotion recognition ability and prosodic discrimination ability, and that the association between vocal emotion recognition and musical aptitude would be stronger than that with musical training, can therefore be confirmed.

Mediation analysis

As shown in Fig. 1, we observed a significant association between musical aptitude and prosodic discrimination ability ($B = 0.32$, $SE = 0.03$, $\beta = 0.62$, $p < 0.001$; path A), as well as a significant association between prosodic discrimination ability and vocal emotion recognition ability ($B = 0.76$, $SE = 0.37$, $\beta = 0.22$, $p = 0.040$; path B). The effect of musical aptitude ($B = 0.39$, $SE = 0.15$, $\beta = 0.22$, $p = 0.010$; total effect C) disappeared when prosodic discrimination ability was included into the model ($B = 0.15$, $SE = 0.19$, $\beta = 0.08$, $p = 0.437$; direct effect C'). This result corresponds to a full mediation of the relationship between musical aptitude and vocal emotion recognition by prosodic discrimination ability (indirect effect = 0.14, 95% CI [0.01, 0.26]), thus confirming hypothesis 4.

We should note that we performed the same mediation analysis while (1) controlling for musical expertise so as to account for the possible influence of music training, and (2) excluding ERAM emotions that were present in the prosodic discrimination test (namely happiness and sadness), with no change in results.

In an additional analysis, we examined the presence of mediation for the two subcomponents of the prosody test. While it might have been expected that a mediation via the emotional test component would be more pronounced, the mediation effects of the emotional test component disappeared when the parametric test component was controlled for. More specifically, while there was a significant association between musical aptitude and prosodic discrimination skills ($B = 0.49$, $SE = 0.07$, $\beta = 0.32$, $p < 0.001$; path A), there was no significant association between the emotional test component and vocal emotion recognition ability ($B = 1.19$, $SE = 0.74$, $\beta = 0.19$, $p = 0.110$; path B) and no mediation effect (indirect effect = 0.11, 95% CI [- 0.02, 0.27]).

These results (see Table S4 for more details), underscore the essential role of both the ability to detect parametric changes and emotional changes in the voice for vocal emotion recognition.

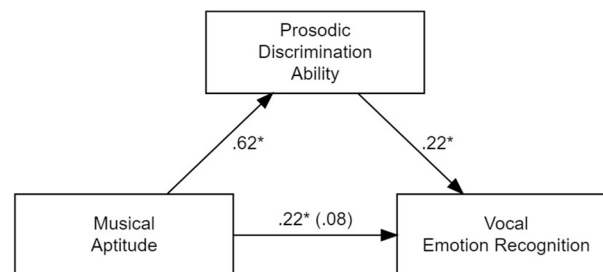


Figure 1. Illustration of the association between musical aptitude and vocal emotion recognition, mediated by prosodic discrimination ability. Standardized scores (β -values) are reported.

Discussion

In Study 3, we examined the relationships between musical aptitude, musical expertise, prosodic discrimination skills, and emotion recognition in a sample of 136 participants with varying levels of musical training. As hypothesized, our results revealed a significant association between vocal emotion recognition and musical aptitude, exceeding the strength of the association with musical expertise. The association between musical aptitude and emotion recognition was fully mediated by individuals' prosodic discrimination skills.

Vocal emotion recognition, musical aptitude, and musical expertise

A key finding of the present study was the robust association between vocal emotion recognition and musical aptitude, which was stronger than with musical expertise. This finding is consistent with recent research highlighting the importance of musical perceptual abilities for speech processing and vocal emotion recognition (e.g.,^{9,35,36,63}), as opposed to the previous focus on musicianship (e.g.,³). Indeed, we found that individuals with above-average music perception abilities but no prior musical training (also referred to as “musical sleepers”³⁴), showed advantages in vocal emotion recognition compared to individuals with low aptitude and training in our sample. In contrast to other studies (e.g.,^{3,64}), we did not find a significant association between vocal emotion recognition and musical expertise.

The mediating role of prosodic discrimination skills

In an attempt to explain the mechanism underlying the relationship between musical ability and vocal emotion recognition, our study showed a mediating role of prosodic discrimination abilities. Consistent with the meta-analysis by Jansen et al.⁹, we observed a stronger correlation between prosodic discrimination skills and musical aptitude, compared to musical expertise. As prosodic discrimination can be distinguished from very basic perceptual abilities (Study 2), the ingredient of an advantage in vocal emotion recognition seems to lie in the enhanced perception of nuances in speech prosody that carry emotional information.

Our exploration of specific subcomponents of musical aptitude highlights the importance of melody and timbre discrimination for vocal emotion recognition, followed by rhythm, tempo, and pitch discrimination. No associations with vocal emotion recognition were observed for the tuning and accent subtests. These results are roughly in line with our expectations that perceptual abilities related to melody, pitch, timbre, tempo, and rhythmic accents should play a particular important role in vocal emotion recognition^{10,37,38}.

The strong correlation between the melody subtest and vocal emotion recognition is reasonable given that emotional messages in both music and speech are conveyed through melodic patterns, such as falling pitch patterns to express sadness^{10,65}. As emotion recognition in both modalities is not only based on individual sounds but rather on their progression within a musical or spoken melody, the minor role of pitch and intonation in our study is not particularly surprising. The association between vocal emotion recognition and the timbre subtest can be explained by the fact that in vocal emotion expression, timbre-like qualities such as voice tremor, shimmer, and voice roughness convey important information about emotional states^{66–68}.

In contrast, musical tempo and rhythm discrimination were only marginally associated with vocal emotion recognition, while no correlation emerged for the accent subtest of the PROMS. This may be due to the different ways in which accentuation is achieved in the accent subtest compared to accentuation in speech, since the latter involves not only changes in loudness but also changes in pitch⁶⁹.

Similarly, prosodic discrimination skills were predominantly correlated with PROMS melody perception and less with rhythm and accent perception. This seems to corroborate one finding of the meta-analysis conducted by Jansen et al.⁹ which showed that musical ability in general (expertise and aptitude) was strongly associated with prosody perception when presented in terms of pitch changes, but less so when presented in terms of timing changes. Taken together, these findings suggest that music perception in the domain of rhythm and accent may be less relevant to prosody perception and vocal emotion recognition than are skills in the area of melody and pitch perception. This interpretation does not stand in contrast to our finding that PROMS-tempo was strongly correlated with prosodic discrimination ability since, in speech, general pace can be clearly distinguished from rhythm and duration of speech elements (such as prosodic phrasing, syllable duration, e.g.,⁷⁰).

Direction of effects

In the present work, we tested whether prosodic discrimination skills mediate the association that has previously been found between musical ability and vocal emotion recognition. In line with prior studies (e.g.,^{3,36}), we considered musical ability as an independent factor predicting speech perception and vocal emotion recognition. It should be noted, however, that the direction of effects could move in the opposite direction with advantages in vocal emotion recognition promoting musicality. To our knowledge, the literature has not yet articulated a model that moves from vocal emotion recognition to music perception skills. Although this is an interesting possibility to consider in future research, our aim here was merely to elucidate the role of prosodic discrimination skills in the musicality-to-emotion recognition association. Furthermore, even if the direction of effects were going into the opposite direction, the association between predictor and outcome would still have to be explained, and prosodic discrimination skills would again seem an obvious mediating mechanism to consider.

Implications and future directions

The main finding of this study is that the enhanced vocal emotion recognition found in musical individuals arises from their ability to detect subtle changes in speech prosody, which is consistent with the concept of shared emotional codes across auditory channels (e.g.,¹⁵), musicians' advantages in speech perception²⁹, and the overlapping cognitive and neural mechanisms involved in music and vocal emotion processing (e.g.,^{71,72}). Although our predictions were mostly accurate, most effects were relatively small.

This may be due to the complexity of vocal emotion recognition, which involves multiple stages from sensory processing, the integration of emotionally meaningful cues, to the formation of evaluative judgments⁸. Our measures of prosodic discrimination skills and musical aptitude primarily relate to the first stage of emotion recognition, namely the perception and analysis of speech signals. On the other hand, research indicates that musical activities also affect social skills such as pro-social behavior⁷³ and empathy^{74,75}. Although not assessed here, these factors may influence vocal emotion recognition, especially the *interpretation* rather than the *perception* stage. Our study found a modest but significant correlation between musical expertise and visual emotion recognition, suggesting potential cross-modal effects. Future research could explore the specifics of musical training, such as whether individual and group musical activities have different influences on the perception and interpretation of vocal cues.

From a research methodology perspective, our findings further highlight the limitation of inferring musical ability from musicianship status. Although musical abilities tend to be more prevalent in musically trained than in musically untrained individuals, especially those required for active music making, non-musicians can have perceptual musical skills that are on par with those of musicians. In turn, there are appreciable individual differences in music perception skills among musicians⁵⁹ which are occluded by grouping all musicians in a single category. The noise in the data created by such classification biases may help explain the inconsistency of findings regarding the effects of musical expertise or musicianship on vocal emotion recognition (e.g.,^{76,77}). As a practical recommendation for future research, we encourage the direct assessment of musical ability (e.g.,⁵³).

Strengths and limitations

The studies' strengths lie in using comprehensive test instruments, including the development of a novel instrument for measuring prosodic discrimination ability, the integration of various subcomponents of musical aptitude into the PROMS-S, and the assessment of vocal, visual, and audiovisual emotion recognition ability using the ERAM.

A limitation is that, although several studies have shown that music perception studies conducted in the laboratory and online provide similar findings^{34,78}, we cannot rule out that completing the tasks in a home environment might have introduced a certain degree of noise into the data. Another limitation is that, unlike previous studies that examined extreme groups (non-musicians vs. professional musicians), our sample included individuals ranging from non-musicians to amateur musicians, with few professional musicians. In addition, the low reliability of the ERAM may have led to some attenuation of the reported correlations⁷⁹.

Finally, it is possible that factors not assessed in the current studies may play a role in the association between musical ability and vocal emotion recognition. One example is emotional intelligence⁷⁷, another is personality traits, such as openness or empathy (e.g.,⁸⁰).

Conclusion

The present research makes two main contributions to the literature: First, it introduces a new test instrument for assessing prosodic discrimination ability; second, it sheds light on the associations between musical aptitude, musical expertise, prosodic discrimination ability, and emotion recognition ability. In Studies 1 and 2, we created a prosodic discrimination test and established its reliability and validity in assessing individuals' ability to discriminate prosodic features in vocal expressions. The mediation found in Study 3 suggests that individuals with higher musical aptitude have an enhanced ability to perceive and discriminate prosodic features that carry emotional information in vocal expressions, ultimately leading to an advantage in the recognition of emotion conveyed by the voice.

Data availability

The datasets used in all three studies, as well as the stimuli for the prosody test, are available through the Open Science Framework (<https://osf.io/98f6z/>).

Received: 22 October 2023; Accepted: 4 July 2024

Published online: 16 July 2024

References

- Martins, M., Pinheiro, A. P. & Lima, C. F. Does music training improve emotion recognition abilities? A critical review. *Emot. Rev.* **13**, 199–210. <https://doi.org/10.1177/17540739211022035> (2021).
- Nussbaum, C. & Schweinberger, S. R. Links between musicality and vocal emotion perception. *Emot. Rev.* **13**, 211–224. <https://doi.org/10.1177/17540739211022803> (2021).
- Lima, C. F. & Castro, S. L. Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion* **11**, 1021–1031. <https://doi.org/10.1037/a0024521> (2011).
- Thompson, W. F., Schellenberg, E. G. & Husain, G. Decoding speech prosody: Do music lessons help?. *Emotion* **4**, 46–64. <https://doi.org/10.1037/1528-3542.4.1.46> (2004).
- Parsons, C. E. *et al.* Music training and empathy positively impact adults' sensitivity to infant distress. *Front. Psychol.* **5**, 1440. <https://doi.org/10.3389/fpsyg.2014.01440> (2014).
- Pralus, A. *et al.* Emotional prosody in congenital amusia: Impaired and spared processes. *Neuropsychologia* **134**, 107234. <https://doi.org/10.1016/j.neuropsychologia.2019.107234> (2019).
- Lima, C. F. *et al.* Impaired socio-emotional processing in a developmental music disorder. *Sci. Rep.* **6**, 34911. <https://doi.org/10.1038/srep34911> (2016).
- Schirmer, A. & Kotz, S. A. Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* **10**, 24–30. <https://doi.org/10.1016/j.tics.2005.11.009> (2006).
- Jansen, N., Harding, E. E., Loerts, H., Başkent, D. & Lowie, W. The relation between musical abilities and speech prosody perception: A meta-analysis. *J. Phonet.* **101**, 101278. <https://doi.org/10.1016/j.wocn.2023.101278> (2023).

10. Juslin, P. N. & Laukka, P. Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychol. Bull.* **129**, 770–814. <https://doi.org/10.1037/0033-2909.129.5.770> (2003).
11. Grandjean, D., Bänziger, T. & Scherer, K. R. Intonation as an interface between language and affect. in *Understanding Emotions*, Vol. 156, 235–247 (Elsevier, 2006).
12. Peretz, I., Vuvan, D., Lagrois, M. -É. & Armony, J. L. Neural overlap in processing music and speech. *Philos. Trans. R. Soc. Lond. B* **370**, 20140090. <https://doi.org/10.1098/rstb.2014.0090> (2015).
13. Patel, A. D. Language, music, syntax and the brain. *Nat. Neurosci.* **6**, 674–681 (2003).
14. Juslin, P. N. & Timmers, R. Expression and communication of emotion in music performance. in *Handbook of Music and Emotion: Theory, Research, Applications*, 453–489 (Oxford University Press, 2010).
15. Juslin, P. N. & Laukka, P. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *J. New Music Res.* **33**, 217–238. <https://doi.org/10.1080/0929821042000317813> (2004).
16. Bowling, D. L. A vocal basis for the affective character of musical mode in melody. *Front. Psychol.* **4**, 464. <https://doi.org/10.3389/fpsyg.2013.00464> (2013).
17. Curtis, M. E. & Bharucha, J. J. The minor third communicates sadness in speech, mirroring its use in music. *Emotion* **10**, 335–348. <https://doi.org/10.1037/a0017928> (2010).
18. Vidas, D., Dingle, G. A. & Nelson, N. L. Children's recognition of emotion in music and speech. *Music Sci.* <https://doi.org/10.1177/2059204318762650> (2018).
19. Werner, S. & Keller, E. Prosodic aspects of speech. In *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, the Future Challenges* (ed. Keller, E.) 23–40 (Wiley, 1994).
20. Moreno, S. *et al.* Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cereb. Cortex* **19**, 712–723. <https://doi.org/10.1093/cercor/bhn120> (2009).
21. Chobert, J., François, C., Velay, J.-L. & Besson, M. Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cereb. Cortex* **24**, 956–967. <https://doi.org/10.1093/cercor/bhs377> (2012).
22. François, C., Chobert, J., Besson, M. & Schön, D. Music training for the development of speech segmentation. *Cereb. Cortex* **23**, 2038–2043. <https://doi.org/10.1093/cercor/bhs180> (2013).
23. Nan, Y. *et al.* Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proc. Natl. Acad. Sci. USA* **115**, E6630–E6639. <https://doi.org/10.1073/pnas.1808412115> (2018).
24. Degé, F. & Schwarzer, G. The effect of a music program on phonological awareness in preschoolers. *Front. Psychol.* **2**, 124. <https://doi.org/10.3389/fpsyg.2011.00124> (2011).
25. Vidal, M. M., Lousada, M. & Vigário, M. Music effects on phonological awareness development in 3-year-old children. *Appl. Psycholinguist.* **41**, 299–318. <https://doi.org/10.1017/s0142716419000535> (2020).
26. Schön, D., Magne, C. & Besson, M. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* **41**, 341–349. <https://doi.org/10.1111/1469-8986.00172.x> (2004).
27. Chartrand, J.-P. & Belin, P. Superior voice timbre processing in musicians. *Neurosci. Lett.* **405**, 164–167. <https://doi.org/10.1016/j.neulet.2006.06.053> (2006).
28. Parbery-Clark, A., Skoe, E., Lam, C. & Kraus, N. Musician enhancement for speech-in-noise. *Ear Hear.* **30**, 653–661 (2009).
29. Neves, L., Correia, A. I., Castro, S. L., Martins, D. & Lima, C. F. Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. *Neurosci. Biobehav. Rev.* **140**, 104777. <https://doi.org/10.1016/j.neubiorev.2022.104777> (2022).
30. Globerson, E., Amir, N., Golan, O., Kishon-Rabin, L. & Lavidor, M. Psychoacoustic abilities as predictors of vocal emotion recognition. *Atten. Percept. Psychophys.* **75**, 1799–1810. <https://doi.org/10.3758/s13414-013-0518-x> (2013).
31. Most, T. Perception of the prosodic characteristics of spoken language by individuals with hearing loss. *Oxford Handb. Deaf Stud. Lang.* **79**, 79–93 (2015).
32. Bigand, E. & Poulin-Charronnat, B. Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition* **100**, 100–130. <https://doi.org/10.1016/j.cognition.2005.11.007> (2006).
33. Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R. & Ullén, F. Practice does not make perfect: No causal effect of music practice on music ability. *Psychol. Sci.* **25**, 1795–1803. <https://doi.org/10.1177/0956797614541990> (2014).
34. Law, L. N. C. & Zentner, M. R. Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS ONE* **7**, e52508. <https://doi.org/10.1371/journal.pone.0052508> (2012).
35. Mankel, K. & Bidelman, G. M. Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proc. Natl. Acad. Sci. USA* **115**, 13129–13134. <https://doi.org/10.1073/pnas.1811793115> (2018).
36. Correia, A. I. *et al.* Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion* **22**, 894–906. <https://doi.org/10.1037/emo0000770> (2022).
37. Juslin, P. N. & Scherer, K. Vocal expression of affect. In *The New Handbook of Methods in Nonverbal Behavior Research* (eds Harigan, J. A. *et al.*) 65–135 (Oxford University Press, 2005).
38. Patel, A. D. *Music, Language, and the Brain* (Oxford University Press, 2008).
39. Laukka, P. *et al.* Investigating individual differences in emotion recognition ability using the ERAM test. *Acta Psychol.* **220**, 103422. <https://doi.org/10.1016/j.actpsy.2021.103422> (2021).
40. Bänziger, T., Grandjean, D. & Scherer, K. R. Emotion recognition from expressions in face, voice, and body: The multimodal emotion recognition test (MERT). *Emotion* **9**, 691–704. <https://doi.org/10.1037/a0017088> (2009).
41. Srinivasan, R. J. & Massaro, D. W. Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Lang. Speech* **46**, 1–22 (2003).
42. Patel, A. D., Foxtton, J. M. & Griffiths, T. D. Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain Cogn.* **59**, 310–313. <https://doi.org/10.1016/j.bandc.2004.10.003> (2005).
43. Leitman, D. I. *et al.* Sensory contributions to impaired prosodic processing in schizophrenia. *Biol. Psychiatry* **58**, 56–61. <https://doi.org/10.1016/j.biopsych.2005.02.034> (2005).
44. Arias, P., Rachman, L., Lind, A. & Aucouturier, J. J. *IRCAM Neutral-Content Voices (Audio Corpus)*. <https://archive.org/details/NeutralContentSaidEmotionallyDataBaseUploadedVersion> (2015).
45. Rachman, L. *et al.* DAVID: An open-source platform for real-time transformation of infra-segmental emotional cues in running speech. *Behav. Res. Methods* **50**, 323–343. <https://doi.org/10.3758/s13428-017-0873-y> (2018).
46. Audacity Team. *Audacity(R): Free Audio Editor and Recorder* (2021).
47. Limesurvey GmbH. *LimeSurvey: An Open Source Survey Tool*. <http://www.limesurvey.org> (2022).
48. Mamassian, P. Confidence forced-choice and other metaperceptual tasks. *Perception* **49**, 616–635. <https://doi.org/10.1177/030106620928010> (2020).
49. Hautus, M., Macmillan, N. A. & Creelman, C. D. *Detection Theory: A User's Guide* (Routledge Taylor & Francis Group, 2021).
50. Livingston, S. A. *Item Analysis. In Handbook of Test Development* 435–456 (Routledge, 2011).
51. Vokey, J. R. Single-step simple ROC curve fitting via PCA. *Can. J. Exp. Psychol.* **70**, 301–305. <https://doi.org/10.1037/cep0000095> (2016).
52. Aujla, H. do': Sensitivity at the optimal criterion location. *Behav. Res.* **55**, 2532–2558. <https://doi.org/10.3758/s13428-022-01913-5> (2022).

53. Strauss, H., Reiche, S., Dick, M. & Zentner, M. R. Online assessment of musical ability in 10 minutes: Development and validation of the Micro-PROMS. *Behav. Res.* <https://doi.org/10.3758/s13428-023-02130-4> (2023).
54. Soranzo, A. & Grassi, M. Psychoacoustics: A comprehensive MATLAB toolbox for auditory testing. *Front. Psychol.* **5**, 712. <https://doi.org/10.3389/fpsyg.2014.00712> (2014).
55. Hodson, G. Construct jangle or construct mangle? Thinking straight about (nonredundant) psychological constructs. *J. Theor. Soc. Psychol.* **5**, 576–590. <https://doi.org/10.1002/jts5.120> (2021).
56. Leek, M. R. Adaptive procedures in psychophysical research. *Percept. Psychophys.* **63**, 1279–1292. <https://doi.org/10.3758/BF03194543> (2001).
57. Bänziger, T., Mortillaro, M. & Scherer, K. R. Introducing the Geneva multimodal expression corpus for experimental research on emotion perception. *Emotion* **12**, 1161–1179. <https://doi.org/10.1037/a0025827> (2012).
58. Kuder, G. F. & Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika* **2**, 151–160. <https://doi.org/10.1007/BF02288391> (1937).
59. Zentner, M. R. & Strauss, H. Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Ann. N. Y. Acad. Sci.* **1400**, 33–45. <https://doi.org/10.1111/nyas.13410> (2017).
60. Hayes, A. F. *The PROCESS macro for SPSS and SAS (version 2.13)[Software]* (2013).
61. Preacher, K. J. & Hayes, A. F. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav. Res. Method. Instrum. Comput.* **36**, 717–731. <https://doi.org/10.3758/bf03206553> (2004).
62. Schoemann, A. M., Boulton, A. J. & Short, S. D. Determining power and sample size for simple and complex mediation models. *Soc. Psychol. Pers. Sci.* **8**, 379–386. <https://doi.org/10.1177/1948550617715068> (2017).
63. Swaminathan, S. & Schellenberg, E. G. Musical competence and phoneme perception in a foreign language. *Psychon. Bull. Rev.* **24**, 1929–1934. <https://doi.org/10.3758/s13423-017-1244-5> (2017).
64. Twaite, J. T. *Examining Relationships Between Basic Emotion Perception and Musical Training in the Prosodic, Facial, and Lexical Channels of Communication and in Music*. Dissertation. City University of New York (2016).
65. Scharinger, M., Knoop, C. A., Wagner, V. & Menninghaus, W. Neural processing of poems and songs is based on melodic properties. *NeuroImage* **257**, 119310. <https://doi.org/10.1016/j.neuroimage.2022.119310> (2022).
66. Tursunov, A., Kwon, S. & Pang, H.-S. Discriminating emotions in the valence dimension from speech using timbre features. *Appl. Sci.* **9**, 2470. <https://doi.org/10.3390/app9122470> (2019).
67. Johnstone, T. & Scherer, K. R. The effects of emotions on voice quality. in *Proceedings of the 14th International Congress of Phonetic Sciences*, 2029–2032.
68. Laukkanen, A.-M., Vilkmann, E., Alku, P. & Oksanen, H. On the perception of emotions in speech: The role of voice quality. *Logop* **22**, 157–168. <https://doi.org/10.3109/14015439709075330> (1997).
69. Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G. & Scherer, K. R. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *J. Acoust. Soc. Am.* **78**, 435–444. <https://doi.org/10.1121/1.392466> (1985).
70. Sares, A. G., Foster, N. E. V., Allen, K. & Hyde, K. L. Pitch and time processing in speech and tones: The effects of musical training and attention. *J. Speech Lang. Hear. Res.* **61**, 496–509. https://doi.org/10.1044/2017_JSLHR-S-17-0207 (2018).
71. Escoffier, N., Zhong, J., Schirmer, A. & Qiu, A. Emotional expressions in voice and music: Same code, same effect?. *Hum. Brain Mapp.* **34**, 1796–1810. <https://doi.org/10.1002/hbm.22029> (2013).
72. Fiveash, A., Bedoin, N., Gordon, R. L. & Tillmann, B. Processing rhythm in speech and music: Shared mechanisms and implications for developmental speech and language disorders. *Neuropsychology* **35**, 771–791. <https://doi.org/10.1037/neu0000766> (2021).
73. Kirschner, S. & Tomasello, M. Joint music making promotes prosocial behavior in 4-year-old children. *Evol. Hum. Behav.* **31**, 354–364. <https://doi.org/10.1016/j.evolhumbehav.2010.04.004> (2010).
74. Cho, E. The relationship between small music ensemble experience and empathy skill: A survey study. *Psychol. Music* **49**, 600–614. <https://doi.org/10.1177/0305735619887226> (2021).
75. Rabinowitch, T.-C., Cross, I. & Burnard, P. Long-term musical group interaction has a positive influence on empathy in children. *Psychol. Music* **41**, 484–498. <https://doi.org/10.1177/0305735612440609> (2013).
76. Dibben, N., Coutinho, E., Vilar, J. A. & Estévez-Pérez, G. Do individual differences influence moment-by-moment reports of emotion perceived in music and speech prosody?. *Front. Behav. Neurosci.* **12**, 184. <https://doi.org/10.3389/fnbeh.2018.00184> (2018).
77. Trimmer, C. G. & Cuddy, L. L. Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion* **8**, 838–849. <https://doi.org/10.1037/a0014080> (2008).
78. Correia, A. I. et al. Can musical ability be tested online?. *Behav. Res.* **54**, 955–969. <https://doi.org/10.3758/s13428-021-01641-2> (2022).
79. Olderbak, S., Riggenmann, O., Wilhelm, O. & Doeblner, P. Reliability generalization of tasks and recommendations for assessing the ability to perceive facial expressions of emotion. *Psychol. Assess.* **33**, 911–926. <https://doi.org/10.1037/pas0001030> (2021).
80. Vuoskoski, J. K. & Thompson, W. F. Who enjoys listening to sad music and why?. *Music Percept.* **29**, 311–317. <https://doi.org/10.1525/mp.2012.29.3.311> (2012).

Author contributions

JV and MZ developed the study idea; JV designed the prosodic discrimination stimuli and conducted studies 1 and 3; FT and JV conducted study 2; JV analyzed and interpreted the data with additional analyses performed by HS; JV and MZ drafted the manuscript and received critical revisions from FT and HS.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-66889-y>.

Correspondence and requests for materials should be addressed to J.V.

Reprints and permissions information is available at www.nature.com/reprints.

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



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/>.

© The Author(s) 2024