Received: 8 March 2023 | Accepted: 24 September 2023

DOI: 10.1111/bjop.12684

ARTICLE

Musicality – Tuned to the melody of vocal emotions

Christine Nussbaum[1,2] | **Annett Schirmer**[1,3] | **Stefan R. Schweinberger**[1,2,4]

[1]Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University, Jena, Germany

[2]Voice Research Unit, Friedrich Schiller University, Jena, Germany

[3]Institute of Psychology, University of Innsbruck, Innsbruck, Austria

[4]Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Correspondence
Christine Nussbaum, Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Am Steiger 3/1, Jena 07743, Germany.
Email: christine.nussbaum@uni-jena.de

Funding information
C.N. has been supported by the German National Academic Foundation ('Studienstiftung des Deutschen Volkes').

Abstract

Musicians outperform non-musicians in vocal emotion perception, likely because of increased sensitivity to acoustic cues, such as fundamental frequency (F0) and timbre. Yet, how musicians make use of these acoustic cues to perceive emotions, and how they might differ from non-musicians, is unclear. To address these points, we created vocal stimuli that conveyed happiness, fear, pleasure or sadness, either in all acoustic cues, or selectively in either F0 or timbre only. We then compared vocal emotion perception performance between professional/semi-professional musicians ($N = 39$) and non-musicians ($N = 38$), all socialized in Western music culture. Compared to non-musicians, musicians classified vocal emotions more accurately. This advantage was seen in the full and F0-modulated conditions, but was absent in the timbre-modulated condition indicating that musicians excel at perceiving the melody (F0), but not the timbre of vocal emotions. Further, F0 seemed more important than timbre for the recognition of all emotional categories. Additional exploratory analyses revealed a link between time-varying F0 perception in music and voices that was independent of musical training. Together, these findings suggest that musicians are particularly tuned to the melody of vocal emotions, presumably due to a natural predisposition to exploit melodic patterns.

KEYWORDS

fundamental frequency (F0), musicality, parameter-specific voice morphing, timbre, vocal emotion perception

BACKGROUND

High levels of musicality are linked to advantages in non-musical domains, such as speech perception and overall cognitive functioning (Elmer et al., 2018; Schellenberg, 2001, 2016). However, while

several decades of systematic research provide robust evidence for a relationship between musical abilities and relatively distant domains, such as language skills (Elmer et al., 2018; Hallam, 2017), links to more closely related domains such as vocal emotion perception are less well established (Martins et al., 2021; Nussbaum & Schweinberger, 2021). Moreover, although accumulating evidence suggests a vocal emotion perception advantage in musicians compared to non-musicians, the underlying mechanisms remain poorly understood, and therefore, are the subject of an important debate. While high-level supramodal processes such as emotional integration and decision making presumably play a role (Lima & Castro, 2011; Trimmer & Cuddy, 2008), the available evidence more consistently points to low-level auditory sensitivity towards musical and vocal cues mediating the advantage in highly trained musicians (Correia et al., 2022). However, it remains unclear how musicians use different vocal cues to infer vocal emotions, and how this might differ from non-musicians. In the present study, we addressed this issue by investigating the degree to which musicians differ in their use of vocal cues that signal vocal emotion. To this end, we manipulated voices to constrain emotional information to specific acoustic cues, which we then presented in an emotion perception task. Thus, we examined how these cues, in isolation and in combination, inform vocal emotion perception in musicians and non-musicians.

What are the acoustic features of emotions and what is shared between music and voice?

Both music and voices convey emotions. In fact, emotional processing measures have identified remarkable overlap between these domains. Psychological overlap has been demonstrated by priming research, as emotional voice and music primes similarly modulate the semantic processing of subsequent positive and negative targets words (Schirmer et al., 2002; Steinbeis & Koelsch, 2011). On a neural level, emotional processing of musical and vocal sounds recruits shared networks (Aubé et al., 2015; Escoffier et al., 2013; Frühholz et al., 2016). A reasonable explanation for these processing parallels highlights acoustic commonalities between musical and vocal emotions: In both domains, emotions are characterized by similar patterns of acoustic cues such as fundamental frequency (F0), amplitude, timing or timbre (Juslin & Laukka, 2003; Scherer, 1995). F0 refers to a sound's lowest harmonic constituent, which we perceive as pitch. As an aside, the perceived pitch corresponds to F0 even in the absence of the lowest harmonic, a phenomenon called the 'missing fundamental' (Smith et al., 1978). In both voices and music, time-varying pitch contour may be more simply described as melody. Timbre refers to a sound's quality that is independent of F0, timing and amplitude. It enables listeners to distinguish, for example, a trumpet from a violin, or one voice from another even when F0, tempo and loudness are identical. Amplitude and timing relate to the loudness and temporal unfolding of sounds, respectively. Importantly, research suggests that the manner in which acoustic cues combine in the context of emotions is shared between music and voice to some degree. Anger, for example, is often characterized by a high pitch, a rough timbre, a large amplitude and fast speech rate, whereas the opposite holds for sadness (Banse & Scherer, 1996). Note, however, that despite the impressive overlap, certain acoustic aspects are unique to either of these channels (such as tonality and harmonic structures in music), which are also reflected in domain-specific neural responses (Escoffier et al., 2013; Juslin & Laukka, 2003).

Research suggests that the different acoustic cues may play different roles in the perception of distinct emotions, albeit their exact roles remain contentious. In early emotional voice perception studies, F0 has been considered the perceptually dominant cue, alongside with amplitude and timing parameters (Banse & Scherer, 1996; Juslin & Laukka, 2003). Recent work, however, suggests that timbre can play a central role in voice processing, and vocal emotion perception in particular (Nussbaum, Schirmer, & Schweinberger, 2022; Nussbaum, von Eiff, et al., 2022; Piazza et al., 2018; von Eiff et al., 2022). In fact, some data imply that both F0 and timbre carry unique information for different emotions (Anikin, 2020; Grichkovtsova et al., 2012). For example, Nussbaum, Schirmer, and Schweinberger (2022) found F0 to

be more important relative to timbre for the recognition of happiness, fear and sadness, whereas timbre seemed relatively more important for the recognition of pleasure. In music, pitch cues, timing and instrumentation have been highlighted as main tools for composers to convey emotional meaning (Juslin & Laukka, 2003; Schutz, 2017). However, the great variety of music styles, instrumentation-dependent acoustic possibilities or constraints, and performers' degrees of freedom make it hard to draw universal conclusions (Schutz, 2017).

How does musicality benefit vocal emotion perception?

Although methodological heterogeneity and limited test power are challenges to existing studies (Martins et al., 2021; Nussbaum & Schweinberger, 2021; Thompson et al., 2004), musicality seems to be associated with a benefit in vocal emotion perception. To explain this benefit, some authors evoked the concept of auditory sensitivity. When compared with non-musicians, musicians are better at perceiving the pitch, timbre and temporal aspects of musical sounds (Kraus & Chandrasekaran, 2010), and it has been argued that this extends to vocal sounds (Chartrand & Belin, 2006; Correia et al., 2022). Yet, exactly how acoustic processing differs between musicians and non-musicians remains elusive. One possibility is that compared to non-musicians, musicians use all acoustic cues more efficiently (e.g. faster, to a greater extent) leading to a *general* improvement of vocal emotion perception. Alternatively, musicality may affect the perception of individual acoustic cues and improve performance in a *cue-specific* way.

Some authors favour a cue-specific benefit and propose a special role of pitch contour (F0). For example, Globerson et al. (2013) identified time-varying pitch perception as a predictor for vocal emotion perception performance. Similarly, pitch is implicated by evidence from participants with amusia, a selective deficit for the processing of musical sounds, despite normal hearing and cognitive abilities (Ayotte et al., 2002; Stewart et al., 2006). In people with amusia, a consistent disadvantage for vocal emotion perception has been reported and linked to problems in pitch discrimination (Lima et al., 2016; Lolli et al., 2015; Pralus et al., 2019; Thompson et al., 2012). However, individuals with amusia represent the tail-end of the musicality spectrum. Their performance does not readily lend itself to inferences about what is special in highly trained musicians. Further, in most studies, amusia was defined based on pitch perception problems only, neglecting the potential influence of other vocal cues (Lagrois & Peretz, 2019). In general, a research focus on pitch may have precluded the potential role of other cues, such as timbre, which has recently been shown to play a significant role in vocal emotional processing (Nussbaum, Schirmer, & Schweinberger, 2022; Nussbaum, von Eiff, et al., 2022). Indeed, research that linked response patterns to different acoustic cues suggests general rather than cue-specific differences between musicians and non-musicians (Lima & Castro, 2011). Thus, a systematic investigation of how musicians and non-musicians use different vocal cues for emotion perception is pending.

Besides auditory sensitivity, high-level supramodal processes have been raised as relevant for explaining performance differences between musicians and non-musicians (Trimmer & Cuddy, 2008). In that vein, musicality has been linked to skills like empathy, emotional differentiation, mind reading and decision making, all of which could foster emotional processing (Clark et al., 2015; Lima & Castro, 2011; Trimmer & Cuddy, 2008). However, a benefit of musicality for emotional processing seems contained within the auditory modality, as it has not been observed for facial or lexical stimuli (Correia et al., 2022; Farmer et al., 2020; Twaite, 2016; Weijkamp & Sadakata, 2017). Further, a comparison of brain responses to vocal emotions between musicians and non-musicians suggests differences at early stages associated with acoustic analysis (Pinheiro et al., 2015; Rigoulot et al., 2015; Strait et al., 2009). Finally, Correia et al. (2022) found that the link between music training and vocal emotion perception was fully mediated by auditory perception skills. Taken together, these findings suggest that the link between musicality and vocal emotion perception is largely acoustic bound.

Methodological challenges and aims of the present study

As mentioned above, some evidence is in line with the proposal that pitch sensitivity explains the superior performance of musicians in vocal emotion perception. However, the neglect of non-pitch cues such as timbre, as well as a reliance on individuals with amusia, makes this evidence inconclusive. Additionally, most of the reported evidence is purely correlational in nature, and therefore, fails to establish a causal link between acoustic cues and emotion perception performance. This situation has recently contributed to an explicit call for more use of voice manipulation tools (Arias et al., 2021). The present study sought to tackle these issues by employing parameter-specific voice morphing. This tool allows, among other things, a resynthesis of vocal stimuli such that they express emotional information through pitch contour or timbre cues only, while rendering the respective other cue uninformative (Kawahara et al., 2008; Kawahara & Skuk, 2019). Here, we manipulated two types of acoustic cues, F0 and timbre, while holding amplitude and timing information constant. We specifically opted for F0 cues because they seem to play an important role for the link between musicality and vocal emotion perception and chose timbre because this cue may hold relevant information for musicians who have to distinguish multiple instrumental or vocal timbres as part of their profession. We were also interested in the relative role of timbre because it is underrepresented in the literature in comparison to F0. Accordingly, we assessed the relative importance of pitch (F0) and timbre for emotional judgements in musicians and non-musicians.

In the present study, we pursued two objectives: The first was a replication of the musicians' advantage for vocal emotional judgements, by recruiting a well-powered sample with highly trained musicians and non-musicians. Second, we assessed how musicians and non-musicians differed in their use of acoustic cues to infer vocal emotions, focusing on the relative importance of F0 versus timbre. Considering the prior work discussed above, we expected that musicians would outperform non-musicians in a condition with full emotion modulation and also when only F0 was informative of emotion. Given the scarcity of data examining timbre, we were also interested in whether the timbre of vocal emotions would also be processed more efficiently by musicians when compared with non-musicians.

METHOD

Participants

In line with previous research comparing vocal emotion perception between musicians and non-musicians (Lima & Castro, 2011), we aimed at a sample size of 40 participants per group. A power analysis using the R-package Superpower (Lakens & Caldwell, 2019) revealed that this sample size would allow the detection of a medium effect ($f = 0.25$) for an interaction between group (musicians, non-musicians) and the stimulus morphing condition (Full, F0 and Timbre) with 80% power.

Data collection took place from June 2021 to May 2022. All participants were fluent German speakers, aged between 18 and 50 years, and provided informed consent before completing the experiment. Data were collected pseudonymized. Participants were compensated with 25€ or with course credit. The experiment was in line with the ethical guidelines of the German Society of Psychology (DGPs) and approved by the local ethics committee of the Friedrich Schiller University Jena (Reg.-Nr. FSV 19/045).

Musicians

We recorded data from 41 (semi-)professional musicians that is individuals with either a music-related academic degree or a non-academic music qualification (details below). The data from two musicians had to be excluded because they omitted >5% trials in the emotion classification task. Thus, data from 39 musicians entered analysis (19 male, 20 female, aged 20–42 years [$M = 29.6$; $SD = 5.64$]). Mean onset age of musical training was 7 years ($SD = 2.53$, 4–17 years). Twenty-four

participants had a music-related academic degree, all others had a non-academic music qualification (i.e. they worked as musicians or won a music competition; for more details see OSF, supplemental tables). Thirty-five participants had studied their instrument for over 10 years, three between 6 and 9 years and one between 4 and 5 years.[1]

Non-musicians

Our recruitment criteria specified that non-musicians had not learned an instrument and did not engage in any musical activities like choir singing during childhood. We recorded data from 40 non-musicians, of which two exceeded the >5% omission criterion. Thus, we analysed data from 38 non-musicians (18 male, 20 female, aged 19–48 years [$M = 30.5$; $SD = 6.54$]). Despite specifying inclusion/exclusion criteria during recruitment, 11 participants later reported having pursued learning an instrument or singing for a short period of time (two reported 2 and three reported 4–5 years of formal musical training[2]; mean age at onset was 16 [$SD = 10.44$, range = 6–30 years]; for details see OSF, Table S1). These participants were retained for data analysis.

Stimuli

Original audio recordings

We selected original audio recordings from a database of vocal actor portrayals provided by Sascha Frühholz, similar to the ones used in Frühholz et al. (2015). For the present study, we used three pseudowords (/molen/, /loman/, /belam/) uttered by eight speakers (four male, four female) with expressions of four emotions (happiness, pleasure, fear and sadness), resulting in a total of 96 original recordings used for morphing. We opted for these four emotions to include two positive and two negative emotions with different degrees of intensity, which was ensured through a prior rating study (reported on OSF).

Voice averaging

Using the Tandem-STRAIGHT software (Kawahara et al., 2008, 2013), we created emotional averages from the four emotions used in the study (see Figure 1) for each speaker and pseudoword. These averages, although not neutral, were assumed to be uninformative and unbiased with respect to the four emotions of interest. We opted for average rather than neutral stimuli because a previous study showed that averages are more suitable for the subsequent generation of voice morphs ensuring that such morphs do not differ systematically in perceived naturalness (Nussbaum et al., 2023).

Parameter-specific voice morphing

To synthesize parameter-specific emotional voice morphs, we created morphing trajectories between each emotion and the emotional average of the same speaker and pseudoword. After manual

[1]Note that the participant reporting 4–5 years of musical training works as a composer and sound engineer and holds an academic music degree, but answered the question about years of formal musical training with regard to formal instrumental lessons only.

[2]Of the three cases of non-musicians who reported 4–5 years of training, one is a hunter who uses a special horn for warning signals, but usually only produces one pitch with it. The other two reported mandatory flute lessons in primary school, which in both cases was several decades ago. After careful consideration of each single case, we decided to keep all participants in the sample.

mapping of time- and frequency anchors at key features of a given utterance pair (e.g. on- and offset of vowels), vocal samples on an emotion/average-continuum were synthesized via weighted interpolation of the originals; for a more detailed description see Kawahara and Skuk (2018). Crucially, Tandem-STRAIGHT allows independent interpolation of five different parameters: (1) F0-contour, (2) timing, (3) spectrum-level, (4) aperiodicity and (5) spectral frequency; the latter three are summarized as timbre.

We created three types of morphed stimuli (see Figure 2). Full-Morphs were stimuli with all Tandem-STRAIGHT parameters taken from the emotional version (corresponding to 100% from the emotion and 0% from average), with the exception of the timing parameter, which was taken from the average (corresponding to 0% emotion and 100% average). F0-Morphs were stimuli with the F0-contour taken from the emotional version, but timbre and timing taken from the average. Timbre-Morphs were stimuli with all timbre parameters taken from the emotional version, but F0 and timing from the average. In addition, all average stimuli were included as a further ambiguous reference category for exploratory purposes. Note that the timing was kept constant across all conditions to allow a pure comparison of F0 versus timbre. In total, this resulted in 8 (speakers) × 3 (pseudowords) × 4 (emotions) × 3 (morphing conditions) + 24 average (8 speakers × 3 pseudowords) = 312 stimuli. For analysis purposes, we collapsed data across speakers and pseudowords.

Using PRAAT (Boersma, 2018), we normalized all stimuli to a root mean square of 70 dB SPL (duration $M = 780$ ms, range 620–967 ms, $SD = 98$ ms). Please refer to OSF, Tables S3 and S4, for a detailed summary of acoustic parameters and some examples of the sound files.

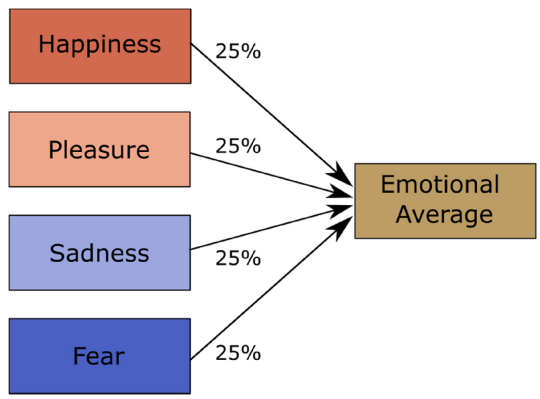

FIGURE 1 Schematic depiction of the voice averaging process.

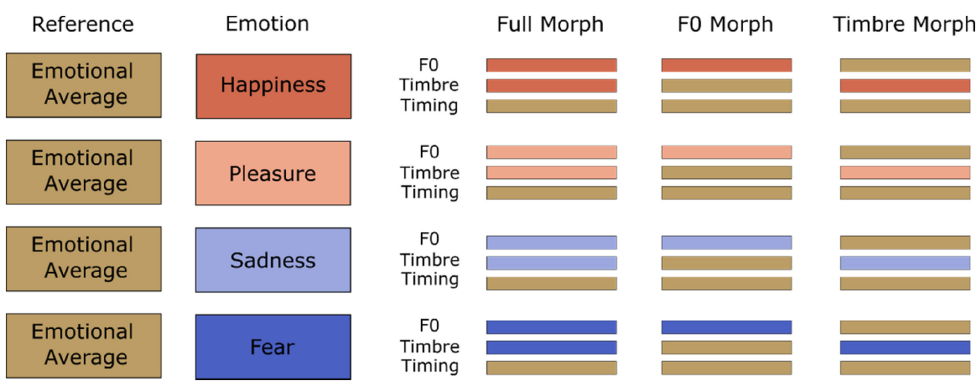

FIGURE 2 Morphing matrix for stimuli with averaged voices as reference.

Design

The study consisted of two sessions: all participants first completed an online session outside the laboratory and were subsequently invited to an EEG session in the laboratory. Here, we only report the results of the online study.

Data were collected online via PsyToolkit (Stoet, 2010, 2017). Participants were required to use a computer with a physical keyboard and headphones, and were asked to ensure a quiet environment for the duration of the study. As browser, we recommended Google Chrome, and excluded Safari for technical reasons. In the beginning, participants entered demographic information, including age, sex, native language, profession and potential hearing impairments, such as tinnitus. Next, participants had the opportunity to adjust their sound settings to a comfortable sound pressure level.

Emotion classification experiment

The participants' task was to classify vocal emotions as happiness, pleasure, fear, or sadness. Each trial started with a green fixation cross presented for 500 ms. Subsequently, a loudspeaker symbol appeared, and the sound was played. After voice offset, a response screen showed the emotion labels and participants could enter their response within a 5000 ms time window starting from voice offset. Participants responded with their left and right index and middle fingers. The mapping of response keys to emotion categories was randomly assigned for each participant, out of four possible key mappings. Emotions of the same valence were always assigned to the same hand and emotions with similar intensity (fear–happiness and sadness–pleasure) were always assigned to the corresponding fingers of both hands (for details, see OSF, Tables S5 and S6). In case of no response (omission error), the final trial slide (500 ms) provided feedback prompting participants to respond faster; otherwise, the screen turned black. Then the next trial started.

At the beginning of the experiment, participants completed eight practice trials with stimuli not used during the actual task. Subsequently, all 312 experimental stimuli were presented once in randomized order across six blocks of 52 trials each. Between blocks, participants could take self-paced breaks. The total duration of the experiment was about 25 min.

Profile of Music Perception Skills (PROMS)

To measure music perception skills beyond self-reports, we adopted the modular version of the Profile of Music Perception Skills (Law & Zentner, 2012; Zentner & Strauss, 2017). We selected the four subtests 'Melody', 'Pitch', 'Timbre' and 'Rhythm', which we considered most informative for the present research. For each subtest, participants completed 18 items, preceded by one practice trial. During each trial, participants heard a reference stimulus twice, followed by a target stimulus. Then, they indicated whether reference and target were the same or different. Although this was a binary decision, the test employs a 5-point Likert scale with the labels 'definitely same', 'maybe same', 'do not know', 'maybe different' and 'definitely different', which we also adopted here. Participants completed the test in about 20 min. One participant encountered technical problems in the 'Melody' subtest, which was therefore repeated several months later to be included in data analysis.

Questionnaires

After the PROMS, participants completed several questionnaires: the German Version of the Autism Quotient Questionnaire, AQ, (Baron-Cohen et al., 2001; Freitag et al., 2007), a 30-item Personality Inventory measuring the Big-Five domains (Rammstedt et al., 2018), the Goldsmiths Musical

Sophistication Index and the Gold-MSI, (Müllensiefen et al., 2014) to assess participants' degree of self-reported musical skills, experiences and engagement. Subsequently, participants reported their socio-economic background and completed the 20-item version of the Positive-Affect-Negative-Affect-Scale, PANAS (Breyer & Bluemke, 2016; Watson et al., 1988). Mean duration of the whole online experiment was about 75 min.

Data analysis

Data were analysed using R Version 4.1.0 (R Core Team, 2020). Response omissions (~1%) were treated as errors and participants with more than 5% of such omissions were excluded from data analysis (see Participants section). Analyses of Variance (ANOVAs) and correlational analyses were performed on data averaged across speaker and pseudoword. Post-hoc tests were Benjamini–Hochberg corrected where appropriate (Benjamini & Hochberg, 1995). We also conducted supplementary trial-level mixed effects modelling and logistic regressions (reported on OSF – supplemental analysis scripts), which largely replicated the results we report below. For important mean values and effect sizes, we provide confidence intervals. The 95% confidence intervals around mean values were calculated based on the standard error and the sample sizes. The 95% confidence intervals around effect sizes (ω^2 and Cohens *d*) were calculated using the R-package 'effectsize' (version 0.4.5.), based on *F*- and *t*-statistics. Concerning the PROMS, we computed a measure that we thought reflected a combination of classification accuracy and certainty. We coded responses from 0 to 1 in 0.25 steps starting with the 'definitely' correct option down two the 'definitely' incorrect option (thus, 'do not know' was always coded with 0.5) and subtracted 0.5 from the final measure. Thus, a positive score indicates that participants were more correct/confident, whereas a negative score indicates more incorrect/uncertain ratings. We then averaged performance across trials for each subtest. Originally, the test authors recommend a d-prime measure which weighs hits and false alarms for response certainty. The results of for such a d-prime measure converge with our own scoring reported here (see OSF, supplemental analyses).

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations and all measures in the study. Pre-processed data, analysis scripts (including documentation of R-package versions) and supplemental materials can be found in the associated OSF repository (https://doi.org/10.17605/OSF.IO/3JKCQ).

Stimulus examples are also provided, but the whole stimulus set cannot be made available for copyright reasons. This study was not preregistered.

RESULTS

Demographic, musicality and personality characteristics of participants

Musicians and non-musicians did not differ in the socioeconomic status assessed via educational level, $X^2(2, N = 77) = 5.21$, $p = .074$, highest academic degree, $X^2(8, N = 77) = 6.40$, $p = .603$ and household income, $X^2(4, N = 77) = 5.66$, $p = .226$ (details on OSF, Table S2). Further, they were comparable in age as well as positive and negative affect (see Table 1 for a summary of participant characteristics assessed via self-report and music performance in the PROMS). For the Big-Five, slightly higher levels of openness and neuroticism were observed in musicians compared to non-musicians. With respect to autistic traits, musicians and non-musicians did not differ in their overall score. However, there were differences in the two subscales proposed by Hoekstra et al. (2008):

TABLE 1 Characteristics of participants – demography, personality and musicality.

	Musicians	Non-musicians					
	***M* (*SD*)**	***M* (*SD*)**	***t***	***df*[a]**	***p***	**Cohens *d* [95%-CI]**	
Age	29.7 (5.6)	30.5 (6.5)	−0.63	72.82	.528	−0.15 [− 0.61, 0.31]	
PANAS							
Positive affect	3.33 (0.66)	3.10 (0.67)	1.51	74.83	.136	0.35 [−0.11, 0.80]	
Negative affect	1.68 (0.47)	1.49 (0.69)	1.39	65.37	.17	0.34 [−0.15, 0.83]	
Big Five							
Openness	**4.11 (0.50)**	**3.81 (0.80)**	1.99	61.77	.050	0.51 [0.00, 1.01]	*
Conscientiousness	3.49 (0.72)	3.76 (0.72)	−1.63	74.96	.108	−0.38 [−0.83, 0.08]	
Extraversion	3.48 (0.66)	3.38 (0.79)	0.61	72.31	.543	0.14 [−0.32, 0.60]	
Agreeableness	3.91 (0.57)	3.75 (0.66)	1.20	72.93	.236	0.28 [−0.18, 0.74]	
Neuroticism	**2.94 (0.66)**	**2.58 (0.82)**	2.10	70.77	.039	0.50 [0.02, 0.97]	*
AQ							
Total	15.64 (5.03)	17.58 (6.41)	−1.47	70.15	.145	−0.35 [−0.82, 0.12]	
Attention to Detail	**5.46 (2.05)**	**4.32 (2.01)**	2.47	74.99	.016	0.57 [0.11, 1.03]	*
Social Communication	**10.18 (4.72)**	**13.26 (6.51)**	2.38	67.38	.020	−0.58 [−1.06, −0.09]	*
Social Skills	**1.44 (1.68)**	**2.61 (2.63)**	−2.32	62.75	.024	−0.59 [−1.09, −0.08]	*
Communication	1.87 (1.63)	2.39 (1.73)	−1.37	74.39	.176	−0.32 [−0.77, 0.14]	
Imagination	**2.13 (1.51)**	**2.87 (1.95)**	−1.86	69.69	.067	−0.45 [−0.92, 0.03]	*t*
Attention Switching	4.74 (1.93)	5.39 (1.92)	−1.48	74.96	.142	−0.34 [−0.80, 0.11]	
Gold-MSI							
General Sophistication	**5.68 (0.50)**	**2.74 (1.07)**	15.38	52.28	<.001	4.25 [3.27, 5.23]	***
Active Engagement	**4.94 (0.82)**	**2.95 (1.19)**	8.50	65.23	<.001	2.11 [1.50, 2.70]	***
Musical Training	**5.94 (0.56)**	**1.71 (0.68)**	29.79	71.75	<.001	7.03 [5.79, 8.27]	***
Emotions	**5.88 (0.74)**	**4.95 (1.32)**	3.79	57.60	<.001	1.00 [0.45, 1.54]	***
Singing Abilities	**5.33 (0.84)**	**2.84 (1.26)**	10.21	64.23	<.001	2.55 [1.89, 3.20]	***
Perceptual Abilities	**6.33 (0.5)**	**4.22 (1.49)**	8.25	45.16	<.001	2.45 [1.68, 3.22]	***
PROMS							
Pitch	**0.27 (0.06)**	**0.18 (0.06)**	6.23	74.97	<.001	0.78 [0.31, 1.25]	***
Melody	**0.23 (0.10)**	**0.07 (0.08)**	9.68	74.95	<.001	−1.96 [1.40, 2.51]	***
Timbre	**0.32 (0.08)**	**0.26 (0.09)**	2.91	73.47	.004	0.44 [−0.02, 0.90]	**
Rhythm	**0.32 (0.08)**	**0.27 (0.08)**	3.35	74.99	.001	0.61 [0.13, 1.08]	**

Note: Descriptive values show mean ratings for the PANAS (Breyer & Bluemke, 2016), the Big-Five Domains (Rammstedt et al., 2018), and the Gold-MSI (Müllensiefen et al., 2014). AQ scores were calculated based on Hoekstra et al. (2008) and Baron-Cohen et al. (2001). * p < .05, ** p < .01, *** p < .001, t .05, p < 0.1. Significance values are provided in each cell in parenthesis.

[a]Note that original degrees of freedom were 75 but were corrected due to unequal variance.

Musicians scored higher than non-musicians on the Attention to Detail subscale, but lower on the Social Communication subscale. Splitting the Social Communication subscale into the four subscales originally proposed by Baron-Cohen et al. (2001), group differences were due to self-reported Social Skills and, although to a lesser degree, to Imagination rather than to Communication or Attention Switching. In the Gold-MSI, musicians scored considerably higher than non-musicians on all subfactors as well as the general musicality score. Further, musicians outperformed non-musicians in all four subtests of the PROMS.

Emotion classification performance

Proportion of correct classifications

The mean proportion of correct responses was submitted to an ANOVA with Emotion (Happiness, Pleasure, Fear and Sadness) and Morph Type (Full, F0 and Timbre) as repeated measures factors and Group (musicians and non-musicians) as a between subject factor. Reference stimuli (emotional averages) were excluded from this analysis. In addition to examining the proportion of correct responses, we also examined unbiased hit rates H_u as outcome measure, as proposed by Wagner (1993). As both approaches yielded identical results with only one exception (reported below), we decided to report the simpler accuracy data here.

Our results included main effects of Group, $F(1, 75) = 5.937$, $p = .017$, $\omega^2 = .06$, 95%-CI [0.00, 0.19], Emotion, $F(3, 225) = 74.18$, $p < .001$, $\omega^2 = .49$, 95%-CI [0.40, 0.56], and Morph Type, $F(2, 150) = 905.25$, $p < .001$, $\omega^2 = .92$, 95%-CI [0.90, 0.94], $\varepsilon_{HF} = .902$. These were qualified by an interaction of Group × Morph Type, $F(2, 150) = 6.10$, $p = .005$, $\omega^2 = .06$, 95%-CI [0.00, 0.14], $\varepsilon_{HF} = .902$ as well as an interaction of Emotion × Morph Type, $F(6, 450) = 26.44$, $p < .001$, $\omega^2 = .25$, 95%-CI [0.18, 0.31], $\varepsilon_{HF} = .904$. The three-way interaction did not reach significance, $F(6, 450) = 0.665$, $p = .663$.[3]

Post-hoc tests revealed that musicians outperformed non-musicians in Full- and F0-morph conditions, whereas there was no difference in the Timbre-morph condition, Full: $|t(69.15)| = 3.35$, $p = .001$, musicians: 77% [75, 79], non-musicians: 72% [69, 74], $d = 0.81$ [0.31, 1.29]; F0: $|t(67.97)| = 2.31$, $p = .023$, musicians: 63% [61, 65], non-musicians: 60% [57, 62], $d = 0.56$ [0.07, 1.04]; Timbre: $|t(74.95)| = 0.30$, $p = .769$, musicians: 41% [39, 43], non-musicians: 41% [39, 43], $d = 0.07$ [−0.38, 0.52], see Figure 3.

Follow-up analyses of the Morph Type effect revealed that performance was best in the Full condition, followed by the F0 and then the Timbre condition (Full: 74% [73, 76], F0: 61% [60, 63], Timbre:

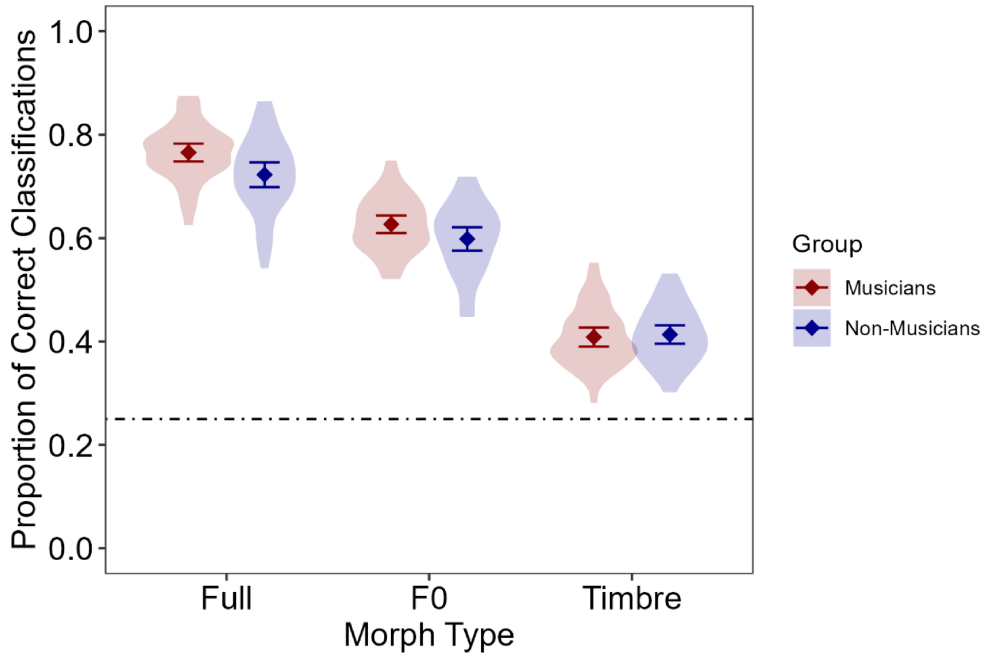

FIGURE 3 Mean proportion of correct responses per Morph Type separately for musicians and non-musicians. Whiskers represent 95% confidence intervals. Violin plots represent variation of individual participants. The dotted line represents guessing rate at .25.

41% [40, 42]; Full vs. F0: $|t(76)| = 20.12$, $p < .001$, $d = 2.31$ [1.88, 2.74], F0 vs. Timbre: $|t(76)| = 22.34$, $p < .001$, $d = 2.56$ [2.10, 3.03], Full vs. Timbre: $|t(76)| = 38.50$, $p < .001$, $d = 4.43$ [3.69, 5.15]). This Morph Type main effect was also found for all emotions separately, all Fs(2, 152) > 116.05, $p < .001$, although it differed slightly between emotions, as suggested by the interaction (see Figure 4, for all post-hoc tests, refer to OSF, supplemental analyses).

To address our specific interest in the relative importance of F0 and Timbre for the different emotions, we calculated the performance difference$_{\text{F0-Tbr}}$ for each emotion separately. Performance difference was largest for Happiness (33% [28, 37]), followed by Fear (23% [19, 26]), Sadness (17% [14, 20]) and Pleasure (9% [6, 12]; all pairwise comparisons $|t$s$(76)| \geq 2.79$, ps$_{\text{corrected}} \leq .006$, ds ≥ 0.32 [0.09, 0.55]). Using unbiased hit rates H_u, the performance difference between Sadness and Pleasure was comparable ($|t(76)| = 1.34$, $p = .184$, $d = 0.15$ [−0.07, 0.38]).

Classification of averaged stimuli and confusion data

In addition to the proportion of correct responses, we calculated confusion data for each Emotion and Morph Type, this time including the averaged stimuli. The response matrices are displayed in Figure 5. A planned analysis of the averaged stimuli revealed that they were most often classified as expressing sadness, followed by pleasure, happiness and fear (sadness vs. pleasure: $|t(76)| = 3.56$, $p < .001$, $d = 0.41$ [0.17, 0.64];

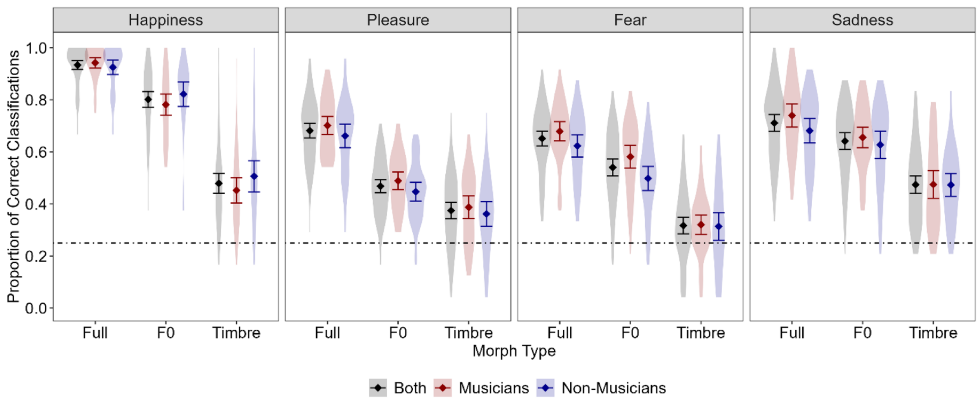

FIGURE 4 Mean proportion of correct responses per Emotion and Morph Type. Whiskers represent 95% confidence intervals. Violin plots represent variation of individual participants. The dotted line represents guessing rate at .25.

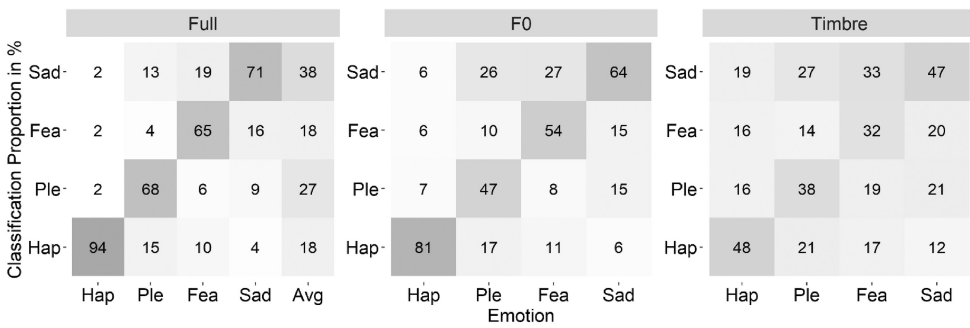

FIGURE 5 Confusion data for each Emotion for the three Morph Types. Numbers represent the proportion of classification responses per Emotion and Morph Type. Hap = happiness, Ple = pleasure, Fea = fear, Sad = sadness, Avg = average.

pleasure vs. happiness: $|t(76)| = 3.40$, $p = .001$, $d = 0.39$ [0.16, 0.63]; happiness vs. fear: $|t(76)| = 0.17$, $p = .867$, $d = 0.02$ [−0.21, 0.24], p-values corrected). There was no significant effect of group. Please refer to our OSF repository, Figures S1 and S2, for a presentation of confusion data separated by group.

Links between musical skills and vocal emotion perception

In a subsequent exploratory analysis, we calculated Spearman correlations between vocal emotion perception performance and both the PROMS music perception performance and the Gold-MSI self-rated musicality. The results are shown in Tables 2 and 3.

Correlations between the PROMS and vocal emotion perception

Of particular interest, we obtained a strong correlation between the overall vocal emotion recognition performance and average PROMS performance. This correlation also emerged in a separate analysis of the control group, $p_s(36) = .48$, $p = .002$, but was non-significant in musicians, $p_s(37) = .22$, $p = .117$, possibly due to reduced variance. Performance in the Full-morph condition correlated with all subtests of the PROMS. Interestingly, there was also a more specific link between the F0 morph condition and the melody subtest, suggesting that both tasks tap into similar abilities. There was no link between the Timbre morph condition and the timbre subtest.

In the next step, we explored these above correlations in more detail to examine a potential role of musical training. Specifically, we calculated partial correlations to control for musical training (Kim, 2015). The correlations between VERAvg and PROMSAvg, $p_s(75) = .41$, $p < .001$, Full-Morphs and Melody, $p_s(75) = .35$, $p = .002$, Full Morphs and Timbre, $p_s(75) = .24$, $p = .036$, and F0-Morphs and Melody, $p_s(75) = .31$, $p = .006$ remained significant. Correlations of Full-morph performance with Pitch and Rhythm turned non-significant when controlling for musical training, $p_s(75) \leq .22$, $p \geq .055$.

TABLE 2 Correlations between the PROMS and vocal emotion perception.

	PROMSAvg	Pitch	Melody	Timbre	Rhythm
VERAvg	**0.44 (<0.001)**	0.22 (0.090)	**0.39 (0.002)**	0.22 (0.084)	**0.35 (0.005)**
Full-Morphs	**0.47 (<0.001)**	**0.28 (0.028)**	**0.44 (<0.001)**	**0.29 (0.023)**	**0.27 (0.028)**
F0-Morphs	**0.35 (0.005)**	0.10 (0.434)	**0.32 (0.011)**	0.15 (0.278)	**0.35 (0.005)**
Timbre-Morphs	0.13 (0.322)	0.11 (0.424)	0.08 (0.523)	0.07 (0.542)	0.13 (0.322)

Note: p-Values (in parenthesis) were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). Significance values are provided in each cell in parenthesis.

Abbreviation: VER, Vocal Emotion Recognition performance.

TABLE 3 Correlations between the Gold-MSI and vocal emotion perception.

	General sophistication	Active engagement	Musical training	Emotions	Singing abilities	Perceptual abilities
VER_{Avg}	**0.30 (0.035)**	0.21 (0.147)	0.21 (0.147)	0.02 (0.865)	**0.31 (0.035)**	**0.28 (0.041)**
Full-Morphs	**0.40 (0.004)**	**0.28 (0.041)**	**0.28 (0.041)**	0.10 (0.555)	**0.41 (0.004)**	**0.34 (0.021)**
F0-Morphs	0.21 (0.147)	0.11 (0.555)	0.12 (0.508)	−0.04 (0.788)	0.23 (0.120)	0.19 (0.168)
Timbre-Morphs	0.08 (0.677)	0.06 (0.741)	0.05 (0.748)	−0.02 (0.865)	0.06 (0.748)	0.09 (0.621)

Note: p-Values (in parenthesis) were adjusted for multiple comparisons using the Benjamini–Hochberg correction (Benjamini & Hochberg, 1995). Significance values are provided in each cell in parenthesis.

Abbreviation: VER, Vocal Emotion Recognition performance.

Correlations between the gold-MSI and vocal emotion perception

There was a correlation between vocal emotion perception performance and self-rated general musical sophistication, with a trend when controlled for musical training, $p_s(75) = .28$, $p = .093$. Further, self-rated singing abilities were linked to increased sensitivity towards vocal emotions, but not when controlled for musical training: $p_s(75) = .23$, $p = .206$. All partial correlations as well as exploratory correlations between the MSI and the PROMS can be found on OSF (Tables S7–S15).

Correlations between personality traits and vocal emotion perception

To rule out that the performance difference between musicians and non-musicians could be attributed to one of the personality traits that differed between groups, we correlated them with averaged vocal emotion performance. The results were non-significant for openness, $p_s(75) = .13$, $p = .269$ but entailed a marginally positive association between neuroticism and vocal emotion perception, $p_s(75) = .22$, $p = .051$. None of the AQ scales correlated significantly with vocal emotion perception performance (all ps $\geq .078$).

DISCUSSION

In this study, we replicated earlier works showing that musicians outperform non-musicians in vocal emotion perception. Further, we investigated the role of different acoustic cues underpinning vocal emotion perception across listener groups and emotional categories. Our findings highlight the special role of pitch contour (F0), that is, the melody of vocal emotions. On the one hand, musicians displayed a specific advantage for this cue. On the other hand, pitch contour seemed to be the perceptually dominant parameter across all emotional categories. In what follows, we will discuss these findings in more detail.

The musicality benefit for vocal emotion perception – a matter of auditory sensitivity?

While the association between high levels of musicality and a benefit in vocal emotion perception has been reported before (Martins et al., 2021; Nussbaum & Schweinberger, 2021), the present study offers an important contribution to this literature. This is because we considered in detail a number of methodological limitations to previous work, including a clear specification of 'musicality', appropriately powered sample sizes, and ensuring comparability with respect to confounding variables such as educational level (Lima & Castro, 2011; Thompson et al., 2004; Trimmer & Cuddy, 2008). In addressing these limitations, the present data offer original and strong evidence for a link between musicality and vocal emotion perception.

Most importantly, our study reveals novel insights into the role of acoustic cues underpinning these benefits. We found that musicians were specifically tuned to the melody of vocal emotions, in that they displayed a small, but reliable cue-specific advantage for pitch contour (F0), but not for timbre. While previous studies reported correlational links between pitch sensitivity and vocal emotion perception (Globerson et al., 2013; Lima & Castro, 2011), we present the first causal evidence for the importance of pitch cues that is based on voice stimuli which were directly acoustically manipulated (Arias et al., 2021).

In line with the general tenor in the literature, our findings suggest that the link between musicality and vocal emotion perception is mediated by low-level auditory sensitivity (Correia et al., 2022; Lima & Castro, 2011; Lolli et al., 2015; Martins et al., 2021) and pitch sensitivity in particular (Globerson et al., 2013). In fact, the link between auditory sensitivity in music and voice perception even holds in the

absence of formal musical education and when correlations are controlled for musical training. These findings converge with data from Correia et al. (2022) who found that the association between music training and vocal emotion perception is fully mediated by auditory and music perception skills. It also fits well with data from individuals with congenital amusia, whose pitch perception deficits predict emotion perception problems (Lima et al., 2016; Lolli et al., 2015; Thompson et al., 2012). Although ours and this latter work do not rule out potential music training effects (Fuller et al., 2018; Good et al., 2017; Thompson et al., 2004), they suggest that differences in auditory sensitivity might prepare some individuals to excel in and enjoy musical activities while also enhancing their vocal socio-affective skills. Nevertheless, one may speculate that if individuals could learn to pay attention to the most informative vocal cues, this could substantially improve their emotion recognition abilities.

Looking at the different subtests of the PROMS allowed us to assess the relevance of specific musical skills for vocal-emotional processing in more detail. Both 'Pitch' and 'Melody' subtests target pitch perception, but 'Pitch' measures pitch discrimination of two static tones, whereas 'Melody' requires the tracking of changes in pitch contour over time (Law & Zentner, 2012; Zentner & Strauss, 2017). Similar to the PROMS 'Pitch' task, the PROMS 'Timbre' task measures the ability to discriminate the timbre of two static tones. However, the PROMS 'Melody' task lacks a timbre equivalent, requiring the tracking of dynamic timbre cues. The 'Rhythm' subtest, by contrast, again requires sensitivity to how acoustic events evolve over time.

In our data, vocal emotion perception was consistently linked to performance in tests that examined time-varying rather than static acoustic processing. Specifically, both the 'Melody' and 'Rhythm' subtests, but not the 'Pitch' and 'Timbre' subtests correlated significantly with overall vocal emotion perception. Thus, for predicting emotion recognition success, tracking acoustic changes over time seems more relevant than representing temporally isolated acoustic features (Juslin & Laukka, 2003). This seems intuitive, as vocal cues are also dynamically evolving over time. Accordingly, we found that the vocal F0 condition correlated with 'Melody', as these tasks share similar demands on the perceptual system, but not with 'Pitch'. Similarly, Globerson et al. (2013) found that vocal prosody recognition could be predicted by the ability to detect dynamic pitch changes, but not by static pitch discrimination. For timbre, the static music task failed to correlate with the vocal timbre condition. Maybe with a music test requiring tracking of timbre features over time a link to vocal timbre perception would become apparent.

Emotional communication in music and voice – same code, same task?

It has been long established that emotions have similar acoustic signatures in voices and music (Juslin & Laukka, 2003). Further, they are perceived in similar ways (Schirmer et al., 2002; Steinbeis & Koelsch, 2011), and processed by shared neural networks (Escoffier et al., 2013; Frühholz et al., 2016; Peretz et al., 2015). The current investigation further strengthens the notion that auditory sensitivity in both domains is linked in listeners. Can we therefore conclude that emotions share the same characteristics and functions in these domains? In traditional models of non-verbal behaviour, emotional prosody has been understood in the context of a sender–receiver perspective, where an emotional message is coded into a signal and the signal is sent with the, perhaps implicit, intent/expectation of being decoded by the receiver (Bänziger et al., 2015; Shariff & Tracy, 2011). Yet, more recently, non-verbal behaviours have been conceptualized more broadly. Accordingly, emotions in voices may not necessarily be a 'message' to another person, but may serve as tool to navigate or influence one's social environment (Schirmer et al., 2022). A fear scream, for example, by sounding unpleasant might serve as a defence mechanism that effectively deters an assailant (Bachorowski & Owren, 1995; Schirmer et al., 2022). These viewpoints do not necessarily exclude each other – auditory emotions are presumably both signals and tools. However, the degree to which different auditory channels serve these functions could differ between music and voices: While vocal emotions result from an agent's current emotional state, musical emotions result perhaps from a more deliberate/explicit communication process. Composers

purposefully translate feelings, states, or intentions into sounds so as to reach an audience. Moreover, music interpreters and performers explicitly reflect on what might be a composer's emotional message as part of their rehearsal work and training. Further, music consumption in Western cultures is predominated by settings with a clear sender/receiver distinction. By contrast, vocal emotions can be found in interactions in which individuals take on more reciprocal roles when behaving non-verbally. Taken together, although vocal and musical emotions share intriguing similarities, they may serve somewhat different functions with the latter being perhaps more intentional in nature.

On a side note, conceptualizing vocal emotions as tools may challenge the ecological validity of explicit emotion categorization tasks, since they do not entirely capture the way vocal emotions are 'used' in daily life (Schirmer et al., 2022). However, it may be expected that musicians can cope better with such an explicit categorizing of emotions, because this approximates their analytic work with music. In the course of practicing a musical piece, emotion categories are often specifically identified, and their expression is expressly pursued. Therefore, future research should probe musicality benefits for vocal emotion perception using implicit measures and brain responses, so as to ascertain that these benefits are not strictly measure dependent (Martins et al., 2022).

The relative importance of pitch contour (F0) and timbre for different emotional categories

In the present data, we found pitch contour (F0) to be more important relative to timbre for successful recognition across all emotional categories. This finding is in line with early work highlighting the importance of pitch cues in vocal emotions (Banse & Scherer, 1996; Juslin & Laukka, 2003). However, performance in the F0 condition, with timbre rendered uninformative, was still worse than in the Full condition, suggesting that timbre carries unique emotional information as well (Grichkovtsova et al., 2012; Nussbaum, Schirmer, & Schweinberger, 2022). This is also reflected in the emotion-specific importance of F0 relative to timbre, which was calculated as the performance difference$_{\text{F0-Tbr}}$ for each emotion separately: The biggest difference between the importance of F0 versus timbre cues was found for happiness, whereas the smallest difference emerged for pleasure and sadness. This finding could be related to studies that highlight the importance of timbre for the perception of sadness (Grichkovtsova et al., 2012) and pleasure (Nussbaum, Schirmer, & Schweinberger, 2022). Minor differences between studies in the relative importance of both acoustic cues for these emotions may be due to their use of different emotional voice databases and the fact that voices can vary substantially in how they are affected by, and communicate, emotions (Spackman et al., 2009). Note that the present work focused on four emotions only. For the future, it would be valuable to expand the present insight to other emotions important to vocal communication, including anger, disgust, surprise and potentially more subtle emotions such as relief, awe or interest (Sauter, 2017).

Constraints on generality and future directions

Although the present study has a number of methodological strengths, including a variety of different speakers and utterances, certain choices in sample and design pose limitations and set directions for further research. One aspect that should be kept in mind is that the present study investigated vocal emotion perception from brief pseudoword stimuli, such that further studies with longer utterances of emotional voices (e.g. sentences or pseudosentences) will be needed to reveal the generality of the present findings. Further, insights gained from the manipulation of acoustic cues by means of voice morphing arguably depend on the degree to which this technology produces ecologically valid stimuli. On the one hand, attention should be dedicated to the naturalness (i.e. human-likeness) of the resulting vocal material (Nussbaum et al., 2023). On the other hand, acoustic features of the emotional averages used to create the parameter-specific voice morphs could introduce uncontrolled emotional

information. Although we assumed that the emotional averages would be ambiguous with respect to the four emotional categories, classification data suggested that they were more often classified as sadness (38%) compared to the other three emotions (18%–27%). While this could be the results of a general sadness-response bias observed in our data (see Figure 5), it bears the risk that averages contained emotionally relevant information. However, we consider it very unlikely that the large performance difference between F0 and timbre conditions would be much affected by the acoustic features of the averaged emotions, because similar results were obtained in a previous study using neutral voices rather than averaged emotions as morphing references (Nussbaum, Schirmer, & Schweinberger, 2022). Finally, we limited our design to the contrast between F0 and timbre only, and therefore did not assess the relative contribution of amplitude and timing.

Regarding the sample, we targeted a population socialized in Western music culture. Additionally, participants were native or fluent German speakers to ensure that the pseudowords used in the study were not perceived as semantically meaningful. Therefore, our findings may not generalize to individuals with a different musical culture or language background (Morrison & Demorest, 2009). Indeed, one would wish to see similar studies conducted with other, more diverse samples.

Further concerning the sample, we note that, despite our best efforts to ensure group comparability, musicians and non-musicians differed in terms of neuroticism and autistic traits. Because these traits did not correlate with vocal emotion perception in the present study, they are unlikely to explain the benefit of musicality. Nevertheless, the differential link between musicality and autistic traits seems worth exploring in more detail, as other studies reported relationships between autistic traits and voice identity perception (Skuk et al., 2019) as well as emotional processing (Yang et al., 2022). While not differing on the total AQ score, musicians seem to score lower on the social communication domain, but higher on the attention to detail domain, when compared with non-musicians. The idea of insular talents including musical aptitude in people with clinical levels of autism is not new (Heaton et al., 1998). Further, autistic traits appear to correlate with pitch perception and absolute pitch in particular (Bonnel et al., 2003; Wenhart et al., 2019). In non-clinical populations, musical skills have been linked to detail-oriented processing (Wenhart & Altenmüller, 2019). However, to date, it is not fully understood how different aspects of autistic traits affect musical aptitude and musical experiences (Sivathasan et al., 2022), which could be worth exploring in the future.

Finally, a particularly interesting comparison for future research would be that between singers and instrumentalists. Our sample was too dominated by instrumentalists to allow for a meaningful analysis of subgroups. Nevertheless, we observed a correlation between self-rated singing abilities and emotion recognition performance. This seems intuitive, since singing provides the form of musical expression that is most closely related to vocal emotions. However, it should be noted that the only study that has compared instrumental versus singing classes suggested that singing could actually interfere with vocal emotion perception (Thompson et al., 2004). This was unexpected, even for the authors, and the degree to which this finding is generated by methodological constraints has been intensely debated (Lima & Castro, 2011; Lolli et al., 2015; Nussbaum & Schweinberger, 2021; Thompson et al., 2004). Of interest in this context, a recent study observed similar brain responses to emotional sounds in singers and instrumentalists (Martins et al., 2022). On balance, at this point, the available literature does not paint a consistent picture concerning the comparison of singers versus instrumentalists, and this issue deserves more systematic investigation.

SUMMARY AND CONCLUSION

Here, we report an advantage for musicians when compared with non-musicians in vocal emotion perception. Moreover, we show, using a novel voice manipulation approach that pitch contour (F0) information plays a more important role than timbre across emotions and listeners and explains the musicality advantage. Further exploratory analyses revealed a link between auditory sensitivity in voices and music, especially for dynamic pitch cues. This link persists in the absence of formal musical training,

suggesting that natural auditory sensitivity, rather than formal music training, drives the benefits of musicality in the context of vocal emotion perception. Future research should expand these findings by comparing different listener subgroups such as singers versus instrumentalists. The possible role of individual differences in personality and autistic traits for the complex interplay between musicality and vocal emotion perception might be another promising path for future exploration.

AUTHOR CONTRIBUTIONS

Christine Nussbaum: Conceptualization; methodology; software; data curation; visualization; formal analysis; writing – original draft. **Annett Schirmer:** Methodology; writing – review and editing; supervision. **Stefan R. Schweinberger:** Conceptualization; writing – review and editing; supervision.

ACKNOWLEDGEMENTS

We thank Bettina Kamchen and Kathrin Rauscher for their support with the data collection and all members of the Voice Research Unit of the Friedrich Schiller University for helpful suggestions on the manuscript. The original voice recordings that served as a basis for creating our stimulus material were provided by Sascha Frühholz. We thank Hannah Strauß for support with the PROMS and Manuel Pöhlmann for help with the OSF repository. We are grateful to all participants of the study. We thank Glenn Schellenberg, Andrey Anikin, and an anonymous reviewer for helpful suggestions on the manuscript. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

C.N. has been supported by the German National Academic Foundation ('Studienstiftung des Deutschen Volkes').

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interests.

DATA AVAILABILITY STATEMENT

Pre-processed data, analysis scripts and supplemental materials can be found in the associated OSF repository (https://doi.org/10.17605/OSF.IO/3JKCQ).

ORCID

Christine Nussbaum https://orcid.org/0000-0003-2718-2898

REFERENCES

Anikin, A. (2020). A moan of pleasure should be breathy: The effect of voice quality on the meaning of human nonverbal vocalizations. *Phonetica*, *77*(5), 327–349. https://doi.org/10.1159/000504855

Arias, P., Rachman, L., Liuni, M., & Aucouturier, J.-J. (2021). Beyond correlation: Acoustic transformation methods for the experimental study of emotional voice and speech. *Emotion Review*, *13*(1), 12–24. https://doi.org/10.1177/1754073920934544

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: Brain responses to music, vocalizations and facial expressions. *Social Cognitive and Affective Neuroscience*, *10*(3), 399–407. https://doi.org/10.1093/scan/nsu067

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

Bachorowski, J. A., & Owren, M. J. (1995). Vocal expression of emotion – Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, *6*(4), 219–224.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Bänziger, T., Hosoya, G., & Scherer, K. R. (2015). Path models of vocal emotion communication. *PLoS One*, *10*(9), e0136675. https://doi.org/10.1371/journal.pone.0136675

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J.-C., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boersma, P. (2018). Praat: Doing phonetics by computer [Computer program]: Version 6.0.46. http://www.praat.org/

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., & Gallun, E. (2003). Enhanced pitch sensitivity in individuals with autism: A signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235. https://doi.org/10.1162/089892903321208169

Breyer, B., & Bluemke, M. (2016). *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel)*. https://doi.org/10.6102/zis242

Chartrand, J.-P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, *405*(3), 164–167. https://doi.org/10.1016/j.neulet.2006.06.053

Clark, C. N., Downey, L. E., & Warren, J. D. (2015). Brain disorders and the biological role of music. *Social Cognitive and Affective Neuroscience*, *10*(3), 444–452. https://doi.org/10.1093/scan/nsu079

Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2022). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*, *22*(5), 894–906. https://doi.org/10.1037/emo0000770

Elmer, S., Dittinger, E., & Besson, M. (2018). One step beyond: Musical expertise and word learning. In *The Oxford handbook of voice perception* (pp. 209–235). Oxford University Press.

Escoffier, N., Zhong, J., Schirmer, A., & Qiu, A. (2013). Emotional expressions in voice and music: Same code, same effect? *Human Brain Mapping*, *34*(8), 1796–1810. https://doi.org/10.1002/hbm.22029

Farmer, E., Jicol, C., & Petrini, K. (2020). Musicianship enhances perception but not feeling of emotion from Others' social interaction through speech prosody. *Music Perception: An Interdisciplinary Journal*, *37*(4), 323–338. https://doi.org/10.1525/MP.2020.37.4.323

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rösler, M., & von Gontard, A. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) – Die Kurzversion AQ-k. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *36*(4), 280–289. https://doi.org/10.1026/1616-3443.36.4.280

Frühholz, S., Klaas, H. S., Patel, S., & Grandjean, D. (2015). Talking in fury: The Cortico-subcortical network underlying angry vocalizations. *Cerebral Cortex*, *25*(9), 2752–2762. https://doi.org/10.1093/cercor/bhu074

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience & Biobehavioral Reviews*, *68*, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

Fuller, C. D., Galvin, J. J., Maat, B., Başkent, D., & Free, R. H. (2018). Comparison of two music training approaches on music and speech perception in Cochlear implant users. *Trends in Hearing*, *22*, 2331216518765379. https://doi.org/10.1177/2331216518765379

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

Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F. A. (2017). Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants. *Ear and Hearing*, *38*(4), 455–464.

Grichkovtsova, I., Morel, M., & Lacheret, A. (2012). The role of voice quality and prosodic contour in affective speech perception. *Speech Communication*, *54*(3), 414–429. https://doi.org/10.1016/j.specom.2011.10.005

Hallam, S. (2017). The impact of making music on aural perception and language skills: A research synthesis. *London Review of Education*, *15*(3), 388–406. https://doi.org/10.18546/Lre.15.3.05

Heaton, P., Hermelin, B., & Pring, L. (1998). Autism and pitch processing: A precursor for savant musical ability? *Music Perception: An Interdisciplinary Journal*, *15*(3), 291–305. https://doi.org/10.2307/40285769

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the autism-Spectrum quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders*, *38*(8), 1555–1566. https://doi.org/10.1007/s10803-008-0538-x

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

Kawahara, H., Morise, M., & Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Kawahara, H., & Skuk, V. G. (2018). Voice morphing. In S. Frühholz, P. Belin, S. Frühholz, P. Belin, & K. R. Scherer (Eds.), *The Oxford handbook of voice perception* (pp. 684–706). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.31

Kawahara, H., & Skuk, V. G. (2019). Voice morphing. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception* (pp. 685–706). Oxford University Press.

Kim, S. (2015). Ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665–674. https://doi.org/10.5351/CSAM.2015.22.6.665

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Lagrois, M.-É., & Peretz, I. (2019). The co-occurrence of pitch and rhythm disorders in congenital amusia. *Cortex*, *113*, 229–238. https://doi.org/10.1016/j.cortex.2018.11.036

Lakens, D., & Caldwell, A. R. (2019). *Simulation-Based Power-Analysis for Factorial ANOVA Designs*. https://doi.org/10.31234/osf.io/baxsf

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), e52508. https://doi.org/10.1371/journal.pone.0052508

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*, 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

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*, 1340. https://doi.org/10.3389/fpsyg.2015.01340

Martins, I., Lima, C. F., & Pinheiro, A. P. (2022). Enhanced salience of musical sounds in singers and instrumentalists. *Cognitive, Affective, & Behavioral Neuroscience*, *22*(5), 1044–1062. https://doi.org/10.3758/s13415-022-01007-x

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

Morrison, S. J., & Demorest, S. M. (2009). Cultural constraints on music perception and cognition. *Progress in Brain Research*, *178*, 67–77. https://doi.org/10.1016/S0079-6123(09)17805-6

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), e89642. https://doi.org/10.1371/journal.pone.0101091

Nussbaum, C., Pöhlmann, M., Kreysa, H., & Schweinberger, S. R. (2023). Perceived naturalness of emotional voice morphs. *Cognition & Emotion*, *37*, 731–747. https://doi.org/10.1080/02699931.2023.2200920

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2022). Contributions of fundamental frequency and timbre to vocal emotion perception and their electrophysiological correlates. *Social Cognitive and Affective Neuroscience*, *17*(12), 1145–1154. https://doi.org/10.1093/scan/nsac033

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

Nussbaum, C., von Eiff, C. I., Skuk, V. G., & Schweinberger, S. R. (2022). Vocal emotion adaptation aftereffects within and across speaker genders: Roles of timbre and fundamental frequency. *Cognition*, *219*, 104967. https://doi.org/10.1016/j.cognition.2021.104967

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1664), 20140090. https://doi.org/10.1098/rstb.2014.0090

Piazza, E. A., Theunissen, F. E., Wessel, D., & Whitney, D. (2018). Rapid adaptation to the timbre of natural sounds. *Scientific Reports*, *8*(1), 13826. https://doi.org/10.1038/s41598-018-32018-9

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*, 107234. https://doi.org/10.1016/j.neuropsychologia.2019.107234

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2018). Validation of the short and extra-short forms of the big five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment*, *36*, 149–161. https://doi.org/10.1027/1015-5759/a000481

R Core Team. (2020). R: A Language and Environment for Statistical Computing. https://www.R-project.org/

Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience*, *290*, 175–184. https://doi.org/10.1016/j.neuroscience.2015.01.033

Sauter, D. A. (2017). The nonverbal communication of positive emotions: An emotion family approach. *Emotion Review*, *9*(3), 222–234. https://doi.org/10.1177/1754073916667236

Schellenberg, E. G. (2001). Music and nonmusical abilities. *Annals of the New York Academy of Sciences*, *930*(1), 355–371. https://doi.org/10.1111/j.1749-6632.2001.tb05744.x

Schellenberg, E. G. (2016). Music training and nonmusical abilities. In *The Oxford handbook of music psychology* (Vol. *2*, pp. 415–429). Oxford University Press.

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*(3), 235–248.

Schirmer, A., Croy, I., & Schweinberger, S. R. (2022). Social touch — A tool rather than a signal. *Current Opinion in Behavioral Sciences*, *44*, 101100. https://doi.org/10.1016/j.cobeha.2021.101100

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*(2), 228–233. https://doi.org/10.1016/S0926-6410(02)00108-8

Schutz, M. (2017). Acoustic constraints and musical consequences: Exploring Composers' use of cues for musical emotion. *Frontiers in Psychology*, *8*, 1402. https://doi.org/10.3389/fpsyg.2017.01402

Shariff, A. F., & Tracy, J. L. (2011). What are emotion expressions for? *Current Directions in Psychological Science*, *20*(6), 395–399. https://doi.org/10.1177/0963721411424739

Sivathasan, S., Philibert-Lignières, G., & Quintin, E.-M. (2022). Individual differences in autism traits, personality, and emotional responsiveness to music in the general population. *Musicae Scientiae*, *26*(3), 538–557. https://doi.org/10.1177/1029864920988160

Skuk, V. G., Palermo, R., Broemer, L., & Schweinberger, S. R. (2019). Autistic traits are linked to individual differences in familiar voice identification. *Journal of Autism and Developmental Disorders*, *49*(7), 2747–2767. https://doi.org/10.1007/s10803-017-3039-y

Smith, J. C., Marsh, J. T., Greenberg, S., & Brown, W. S. (1978). Human auditory frequency-following responses to a missing fundamental. *Science*, *201*(4356), 639–641. https://doi.org/10.1126/science.675250

Spackman, M. P., Brown, B. L., & Otto, S. (2009). Do emotions have distinct vocal profiles? A study of idiographic patterns of expression. *Cognition and Emotion*, *23*(8), 1565–1588. https://doi.org/10.1080/02699930802536268

Steinbeis, N., & Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience*, *23*(3), 604–621. https://doi.org/10.1162/jocn.2009.21383

Stewart, L., von Kriegstein, K., Warren, J. D., & Griffiths, T. D. (2006). Music and the brain: Disorders of musical listening. *Brain: A Journal of Neurology*, *129*(10), 2533–2553. https://doi.org/10.1093/brain/awl171

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

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. *The European Journal of Neuroscience*, *29*(3), 661–668. https://doi.org/10.1111/j.1460-9568.2009.06617.x

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

Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion*, *8*(6), 838–849. https://doi.org/10.1037/a0014080

Twaite, J. (2016). Examining Relationships Between Basic Emotion Perception and Musical Training in the Prosodic, Facial, and Lexical Channels of Communication and in Music.

von Eiff, C. I., Skuk, V. G., Zäske, R., Nussbaum, C., Frühholz, S., Feuer, U., Guntinas-Lichius, O., & Schweinberger, S. R. (2022). Parameter-specific morphing reveals contributions of timbre to the perception of vocal emotions in Cochlear implant users. *Ear and Hearing*, *43*(4), 1178–1188. https://doi.org/10.1097/AUD.0000000000001181

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28. https://doi.org/10.1007/BF00987006

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weijkamp, J., & Sadakata, M. (2017). Attention to affective audio-visual information: Comparison between musicians and non-musicians. *Psychology of Music*, *45*(2), 204–215. https://doi.org/10.1177/0305735616654216

Wenhart, T., & Altenmüller, E. (2019). A tendency towards details? Inconsistent results on auditory and visual local-to-global processing in absolute pitch musicians. *Frontiers in Psychology*, *10*, 31. https://doi.org/10.3389/fpsyg.2019.00031

Wenhart, T., Bethlehem, R. A. I., Baron-Cohen, S., & Altenmüller, E. (2019). Autistic traits, resting-state connectivity, and absolute pitch in professional musicians: Shared and distinct neural features. *Molecular Autism*, *10*, 20. https://doi.org/10.1186/s13229-019-0272-6

Yang, D., Tao, H., Ge, H., Li, Z., Hu, Y., & Meng, J. (2022). Altered processing of social emotions in individuals with autistic traits. *Frontiers in Psychology*, *13*, 746192. https://doi.org/10.3389/fpsyg.2022.746192

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the short-PROMS and the mini-PROMS. *Annals of the New York Academy of Sciences*, *1400*(1), 33–45. https://doi.org/10.1111/nyas.13410

How to cite this article: Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2024). Musicality – Tuned to the melody of vocal emotions. *British Journal of Psychology*, *115*, 206–225. https://doi.org/10.1111/bjop.12684