Vocal Emotion Perception: A Comparison of Singers and Instrumentalists, Amateurs and Professionals

Christine Nussbaum1,2, Jessica Dethloff1,2, Annett Schirmer3,2, and Stefan R. Schweinberger1,2,4

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

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

3Institute of Psychology, University of Innsbruck, Austria

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

Correspondence should be addressed to Christine Nussbaum, Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Leutragraben 1, 07743 Jena, Germany. Tel: +49 (0) 3641 945939, E-Mail: christine.nussbaum@uni-jena.de; or to Stefan R. Schweinberger, Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Am Steiger 3/Haus 1, 07743 Jena, Germany. Tel: +49 (0) 3641 945181, Fax: +49 (0)3641 9 45182, E-Mail: stefan.schweinberger@uni-jena.de. Preprocessed data, analysis scripts and supplemental materials can be found in the associated OSF repository (https://osf.io/ascqx/).

**Word count:** aim 5000

Contents

[*Abstract* 3](#_Toc194687188)

[*Public significance statement* 4](#_Toc194687189)

[1 Introduction: associations between musicality and vocal emotion perception 5](#_Toc194687190)

[1.1 Singers vs. instrumentalists 6](#_Toc194687191)

[1.2 Amateurs vs. professional musicians 7](#_Toc194687192)

[1.3 Rationale, outline etc. 8](#_Toc194687193)

[2 Part I: Comparison of non-professional singers and instrumentalists 9](#_Toc194687194)

[2.1 Hypotheses 9](#_Toc194687195)

[2.2 Method 10](#_Toc194687196)

[2.2.1 Participants 10](#_Toc194687197)

[2.2.2 Stimulus material 11](#_Toc194687198)

[2.2.3 Design 13](#_Toc194687199)

[2.2.4 Data analysis 15](#_Toc194687200)

[2.3 Transparency and openness 16](#_Toc194687201)

[2.4 Results 16](#_Toc194687202)

[2.4.1 Demography, musicality, and personality of participants 16](#_Toc194687203)

[2.4.2 Emotion classification performance 18](#_Toc194687204)

[3 Part II: Comparison of professionals, amateurs and non-musicians 20](#_Toc194687205)

[3.1 Hypotheses 20](#_Toc194687206)

[3.2 Method 21](#_Toc194687207)

[3.3 Results 21](#_Toc194687208)

[3.3.1 Demography, musicality, and personality of participants 21](#_Toc194687209)

[3.3.2 Emotion classification performance 24](#_Toc194687210)

[4 Part III: Correlational analyses 25](#_Toc194687211)

[4.1 Hypotheses 25](#_Toc194687212)

[4.2 Method 25](#_Toc194687213)

[4.3 Results 26](#_Toc194687214)

[5 Discussion 26](#_Toc194687215)

[5.1 Singers vs. instrumentalists 26](#_Toc194687216)

[5.2 Professional musicians vs. amateurs 27](#_Toc194687217)

[5.3 Constraints on generality and future directions 27](#_Toc194687218)

[6 Summary and Conclusion 28](#_Toc194687219)

[7 Acknowledgements 28](#_Toc194687220)

[8 Conflicts of Interests and Funding 28](#_Toc194687221)

[9 Credit Author Statement 28](#_Toc194687222)

[10 Supplementary material 29](#_Toc194687223)

[References 29](#_Toc194687224)

# Abstract

Musicians outperform non-musicians in vocal emotion recognition, presumably due to differences in auditory sensitivity for melodic patterns that carry emotional meaning. However, the current literature is inconclusive regarding differential effects of specific types of musical activity. Given the tight link between expression and perception in emotional communication, we asked if it makes a difference if music is expressed vocally or not, and whether music is performed on an amateur or professional level. Importantly, however, we predicted that vocal emotion recognition would be *unaffected* by the type and amount of musical activity because current evidence argues against a causal role of formal musical education. In the present study, we compared emotion recognition performance of singers (N= 45) vs. instrumentalists (N=43) and professional musicians (N = 40) vs. amateurs (N = 88) vs. non-musicians (N = 38). Using both frequentist and Bayesian inference, we found the predicted null effects for singers vs. instrumentalists, and for professionals vs. amateurs. The pattern for amateurs vs. non-musicians was inconclusive. Across groups, we replicated the consistent link between vocal emotion perception and auditory sensitivity, especially for melodies. Thus, the current work aligns with the perspective that a musicians’ advantage for vocal emotions is tied to natural auditory sensitivity but not the type of musical activities or the amount of formal training.

**Keywords:** vocal emotion perception, singers, instrumentalists, amateurs, parameter-specific voice morphing, musicality

# Public significance statement

* Musicians outperform non-musicians in vocal emotion perception, but it is currently unclear whether this effect is modulated by the specific type of musical activity
* To address this gap, we compared vocal emotion perception in singers vs. instrumentalists and in amateurs vs. professional musicians
* In line with our predictions, we found no difference between either of these subgroups, suggesting that vocal emotion perception is unaffected by the type and amount of musical activity
* Instead, we replicated the link between natural auditory sensitivity and vocal emotion perception

# Introduction: associations between musicality and vocal emotion perception

The human voice is a prime carrier of emotional information. Therefore, adequate perception of vocal emotions is important for everyday social interaction (Laukka et al., 2016; Schirmer et al., 2025). On average, humans can infer emotion from voices well above chance (Banse & Scherer, 1996; Juslin & Laukka, 2003; Scherer, 2018), but this capacity is subject to great individual variability and seems to be linked to differences in **musicality**. It has been shown repeatedly that musicians outperform non-musician in vocal emotion recognition, although the overall effect size can be considered small to moderate (M. Martins et al., 2021; Nussbaum & Schweinberger, 2021; Schellenberg & Lima, 2024). Several works sought to unravel the potential mechanisms underlying this advantage and collectively tend to emphasize the role of **acoustic sensitivity.** Musicians have more fine-grained basic auditory skills compared to non-musicians, such as pitch and rhythm perception, musical memory, or signal-in-noise discrimination (Baldé et al., 2025; Kraus & Chandrasekaran, 2010), which aids vocal emotion recognition. Importantly, this does not seem to depend directly on formal musical training: The association between auditory perception and vocal emotion recognition was observed even in the absence of any formal musical training (Correia et al., 2022; Nussbaum et al., 2024; Vigl et al., 2024). Further, Correia et al. (2022) found that the link between music training and vocal emotion perception was fully mediated via auditory perception skills. The presumably strongest evidence is provided by a recent randomized-controlled study in school children, which found no causal effects of musical training on vocal emotion perception performance (Neves et al., 2025). Thus, there is consensus in the literature that the observed performance difference of musicians and non-musicians is due to variations in acoustic sensitivity rather than the result of formal musical education (Schellenberg & Lima, 2024).

In a previous study, we investigated how musicians’ auditory skills promote vocal emotion perception in more detail, by focusing on different auditory cues that transport emotional meaning (Nussbaum et al., 2024). We employed parameter-specific voice morphing to create vocal stimuli that expressed emotion only through fundamental frequency contour (F0), timbre or both. F0 is linked to dynamic pitch variation (also referred to as voice melody) and timbre is linked to perceived voice quality (i.e. whether it sounds harsh or gentle). Professional musicians outperformed a group of non-musicians when emotions were expressed by F0 and both cues, but not timbre alone. Thus, musicians seem to be specifically proficient at exploiting melodic patterns to infer vocal emotions.

While the available literature paints a fairly consistent picture regarding the link between musicality and vocal emotion perception, a key limitation inherent in most studies targeting group differences is that they treat musicians as one uniform group, whereas they are, in fact, highly heterogeneous. On the one hand, there are quantitative differences regarding levels of expertise. On the other hand, qualitative differences between musicians can relate to a great variety of styles, genres, and forms of expression, within the scope of the Western music system and beyond. A particularly interesting distinction in the context of vocal emotions is the one between **singers and instrumentalists**. Singing is arguably the form of musical expression that is most closely related to vocal emotions (Akkermans et al., 2019; Mithen et al., 2006). Another interesting debate evolves around differences between **professional musicians and amateurs**. Therefore, the present study targeted these subgroups to explore differential patterns in vocal emotion perception. In what follows, we review current insights and outstanding research gaps, cumulating in the rationale for the present study.

## Singers vs. instrumentalists

Singing and playing an instrument are both fundamental forms of musical expression in humans, but they require very different motor skills and, typically, different amounts of formal musical training (Fisher et al., 2020; Krishnan et al., 2018). This is indeed reflected in vocal performance differences (Christiner & Reiterer, 2015). For perception, however, this is less clear. Given the tight links between expression and perception in auditory communication via the voice (Frühholz & Schweinberger, 2021), one could assume that with their supreme expertise in vocal expression, singers might outperform instrumentalists in vocal perception as well. However, this does not seem to be the case. In contrast, the link between vocal pitch production and pitch perception was observed even to a larger degree in instrumentalists than in vocalists (Nikjeh et al., 2009). With regard to vocal emotion perception, evidence is similarly sparse and inconclusive. Several studies observed correlations between vocal emotion perception and singing abilities, either self-rated and objectively measured (Correia et al., 2022; Greenspon & Montanaro, 2023; Nussbaum et al., 2024), but all samples comprised both singers and instrumentalists. On a neural level, I. Martins et al. (2022) found no differences in electrophysiological response to emotional voices between singers and instrumentalists, suggesting similar profiles of auditory processing. Intriguingly, a music-intervention study reported that singing may even interfere with vocal emotional processing, while instrument lessons had a positive effect (Thompson et al., 2004). However, the validity of this finding is limited by an extensive drop-out of participants and a small sample size (Schellenberg & Lima, 2024). Overall, the few data that are available do not provide clearcut evidence for a specific benefit in vocal emotion recognition by singing over playing an instrument. We therefore pursued the null hypothesis of no such differences – but in view of the limitations with previous studies, we recruited a well-powered sample of instrumentalists and singers.

## Amateurs vs. professional musicians

Most musicians start with their formal training in childhood, but when they enter adulthood, they pick different paths: some convert their musical activity into a profession, others pursue another career but keep it as a hobby. Interestingly, these groups seem to display several differences with regard to neurocognitive functioning. While amateurs, unsurprisingly, score lower on musical abilities, they show better cognitive abilities in terms of abstract reasoning than professional musicians (Vincenzi 2022). Amateurs may gain more positive outcomes from their musical activity, perhaps because it provides enrichment in addition to their profession, while coming with less noise and less performative pressure. This also seems to be reflected in general health, which was found to be better in amateurs than professionals (Bonde et al., 2018; Hake et al., 2024; Loveday et al., 2023; Maghiar et al., 2023; Rogenmoser et al., 2018). On a different note, one recent study reported that professionals more often experience a state of flow during their musical activity, which is usually considered very enjoyable (Rakei & Bhattacharya, 2024). However, to the best of our knowledge, there are no insights into differences between amateurs and professionals with regard to vocal emotion perception. This gap is addressed with the present study. Because all our newly recruited singers and instrumentalists were amateurs and we used an established experimental procedure, the present study offered the opportunity to compare findings with our previously recruited groups of professional musicians and non-musicians (Nussbaum et al., 2024).

## Rationale of the present study

This study focuses on the comparison between singers and instrumentalists, and thus zooms into possible differences between specific subgroups while using the almost identical protocol as Nussbaum et al. (2024). Accordingly, we report on the findings in three parts. For Part I, we recruited an original sample of amateur instrumentalists and singers and compared their vocal emotion recognition, their musical perception performance and self-rated musicality. In Part II, we focused on the correlations between these measures in order to replicate the link between auditory sensitivity and vocal emotion recognition reported by previous studies. For Part III, we collapsed all participants recruited here into one group of amateur musicians and compared it to professional musicians and non-musicians from the previous study (Nussbaum et al., 2024).

As mentioned above, we predicted that singers and instrumentalists would perform equally in our vocal emotion recognition task, both for emotions expressed by all available vocal cues, as well as emotions expressed by either F0 or timbre cues in the voice. At first sight, this prediction may seem surprising given the evidence for tight expression-perception-links in vocal communication. However, the above-mentioned literature suggests that this is comparable in singers and instrumentalists. Further, the available evidence consistently argues that the link between musicality and vocal emotion perception is not driven by formal training, but rather by predisposed differences in auditory sensitivity. If this is the case, the form of musical engagement should not make a difference. A similar logic holds for the comparison for professional musicians and amateurs. We therefore predicted that their vocal emotion recognition performance would be comparable as well (for full emotions, as well as the F0 and timbre condition). The whole study and its hypothesis were preregistered (<https://doi.org/10.17605/OSF.IO/76PV5>).

# Part I: Comparison of non-professional singers and instrumentalists

## Hypotheses

Regarding the comparison between singers and instrumentalists, we formulated to following hypotheses:

**H1:** We expect NO difference between singers and instrumentalists in overall vocal emotion recognition performance.

**H2:** We expect NO difference between singers and instrumentalists in vocal emotion recognition performance based on timbre and F0 cues only.

## Method

Note that this is a follow-up to the study reported in Nussbaum et al. (2024). Thus, the stimulus material and the design are almost identical, but we recruited a new sample.

### Participants

According to our preregistered plan, we aimed at a sample size of 40 singers (20 male, 20 female) and 40 instrumentalists (20 male, 20 female), because in our previous study, this sample size allowed us to reveal medium-sized group effects (d = 0.81 for the Full and d = 0.56 for the F0 morphing condition) when we compared professional musicians and non-musicians.

Data were collected in a pseudonymized format from June 2023 to January 2024. All participants were aged between 18 and 54 years and fluent German speakers. Participants provided informed consent before completing the experiment and received compensation in the form of 12.50 € or course credit upon completion. 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).

In total, we collected data from 94 amateur musicians that were divided into singers and instrumentalists. Recruitment criteria specified that participants had to be non-professional musicians (i.e., they held no music-related academic degree or worked professionally as a musician). Singers were required to be currently active in a choir or another singing group but should not play an instrument actively and regularly (i.e., they must not currently be instrumentalists in an orchestra or a band). Instrumentalists, conversely, were required to be currently active in an orchestra or a band, but they should not engage in singing activities actively and regularly (i.e., they must not currently be in a choir or another singing group).

*Singers*

We recorded data from 48 singers, of which three were excluded (N = 2 had > 5 % trials of omission, N = 1 had technical issues during stimulus playback). Thus, data from 45 singers were analyzed (22 female, 22 male, 1 diverse, aged 18 to 53 years [M = 27.02, SD = 8.2]). Mean onset age of musical training was 8 years (SD = 3.08, 5 - 20 years). Mean duration of musical training was 10 years (SD = 1.04, 0 – 25 years). Five participants reported that they never had any formal musical training (for more details see ToDo). Two participants reported that they had occasional tinnitus, but without any subjective impairments in daily life.

*Instrumentalists*

Data from 46 instrumentalists were collected, of which three were excluded. One had technical issues during stimulus playback, one was also active in a choir, one held a master’s degree in music science and was therefore relocated into the group of professional musicians (see Part III). Thus, data from 43 instrumentalists entered analysis (24 female, 18 male, 1 diverse, aged 18 to 54 years [M = 28.51, SD = 10.64]). Mean onset of musical training was 7 years (SD = 2.27, 4 - 14 years). Mean duration of musical training/activity was 14 years (SD = 1.53, 0 – 44 years). Four participants reported that they never had any formal musical training

### Stimulus material

As stimulus material, we used parameter-specific voice morphs that express emotional information either through the fundamental frequency contour only (F0), through timbre only (Tbr) or through a combination of both (Full).

For voice morphing, we selected original audio recordings from a database of vocal actor portrayals, comprised of pseudowords (/molen/, /loman/, /belam/) uttered by eight speakers (four male, four female) with expressions of happiness, pleasure, fear, and sadness. To synthesize the parameter-specific emotional voice morphs, we created morphing trajectories between each emotion and an emotional average of the same speaker and pseudoword, using the Tandem-STRAIGHT software (Kawahara et al., 2013; Kawahara et al., 2008). The averages had been created previously by blending all emotions together and were thus assumed to be uninformative and unbiased with respect to the four emotions of interest. After substantial preprocessing (e.g. manual mapping of time- and frequency anchors in each stimulus), Tandem-STRAIGHT enables voice morphing via weighted interpolation of five independent 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 1**). **Full-Morphs** were stimuli with all parameters taken from the emotional version (corresponding to 100% from the emotion and 0% from average), except for 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 emotion, but timbre and timing were taken from the average. **Timbre-Morphs** were stimuli with all timbre parameters taken from the emotion, but F0 and timing from the average. Note that the timing was kept constant in all conditions to allow a pure comparison of F0 vs. timbre. Furthermore, we kept all average stimuli as a further ambiguous reference category. In total, this resulted in 8 (speakers) x 3 (pseudowords) x 4 (emotions) x 3 (morphing conditions) + 24 average (8 speakers x 3 pseudowords) = 312 stimuli (duration M = 780 ms, range 620 to 967 ms, SD = 98 ms). Using PRAAT (Boersma, 2018), we normalized all stimuli to a root-mean-square of 70 dB SPL.

For a more detailed description of the stimulus creation, see Nussbaum et al. (2024) and Kawahara and Skuk (2019). For a summary of acoustic characteristics, see Tables S3 and S4 on OSF.

**Figure 1**

*Morphing matrix for stimuli with averaged voices as reference*



*Note. Figure reprinted from Nussbaum et al. (2024), Fig 2, page 6*

### Design

Data were collected online via PsyToolkit (Stoet, 2010, 2017), but after completion of the study all participants met with the experimenter for a short personal debriefing. This was done to increase commitment and conscientiousness for the experiment.

Participants were required to ensure a quiet environment for the duration of the study and use a computer with a physical keyboard and headphones. As browser, we recommended Google Chrome, and excluded Safari for technical reasons. Prior to the listening tasks, participants could adjust their sound settings to a comfortable sound pressure level

First, participants entered demographic information, including age, sex, native language, profession, and potential hearing impairments such as tinnitus. They then completed an emotion classification experiment, a test on music perception (the Profile of Music Perception Skills) and several questionnaires on musicality, personality and socioeconomic background. Mean duration of the whole online study was about 75 minutes.

Emotion classification experiment

In the experiment, participants classified vocal emotions as happiness, pleasure, fear, or sadness. Each trial started with a green fixation cross presented for 500 ms. Then the sound was played while a loudspeaker symbol was shown on the screen. Subsequently, a response screen showed the emotion labels and participants could enter their response within a 5000 ms time window starting from voice offset. Responses were entered with the left and right index and middle fingers, with random mapping of response keys to emotion categories for each participant, out of four possible key mappings (see Tables S5 and S6 on OSF). In case of a response omission, the final trial slide (500 ms) prompted participants to respond faster; otherwise, the screen turned back. Then the next trial started.

The 312 stimuli were presented in randomized order in six blocks of 52 trials each, with self-paced breaks in between. Beforehand, participants completed eight practice trials with different stimuli. The experiment was about 25 minutes long. Unfortunately, due to a software bug, randomization was sampled with replacements, so that some stimuli were drawn repeatedly and others were omitted, in contrast to the pre-study, where randomization was sampled without replacement so that each stimulus was drawn exactly once.

Profile of Music Perception Skills (PROMS)

To measure music perception skills, we used a modular version of the Profile of Music Perception Skills (Law & Zentner, 2012; Zentner & Strauss, 2017), comprised of the four subtests „Melody“, „Pitch“, „Timbre”, and „Rhythm“. Participants completed 18 items per subtest, always preceded by one practice trial. 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 via a 5-point Likert scale with the labels “definitely same”, “maybe same”, “don’t know”, “maybe different”, and “definitely different”. The duration of the PROMS was about 20 minutes.

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, Gold-MSI, (Müllensiefen et al., 2014) to assess the participants’ degree of self-reported musical skills, additional questions concerning music experience and musical engagement, their socioeconomic background, and the 20-item version of the Positive-Affect-Negative-Affect-Scale, PANAS (Breyer & Bluemke, 2016; Watson et al., 1988).

### Data analysis

In line with our preregistered plan, we collapsed data across speakers and pseudowords for analysis. Further, data on emotional averages were excluded because they were not relevant for our hypotheses. Response omissions (~1%) were treated as errors and participants with more than 5% of such omissions excluded from data analysis. Analyses of Variance (ANOVAs) and correlational analyses were performed using R Version 4.3.2 (R Core Team, 2020). Post-hoc tests were Benjamini-Hochberg corrected where appropriate (Benjamini & Hochberg, 1995).

We complemented these classical frequentist analyses with a Bayesian approach, which – in contrast to null hypothesis significance testing - allows a quantification of evidence for null findings (Rosenfeld & Olson, 2021). These analyses were conducted in JASP Version 0.19.3 (Quelle) using default priors. We report the Bayes factor (BF10) as an indicator for the likelihood of the null and alternative hypothesis given the observed data. BF10 > 1 indicate larger evidence for the alternative hypothesis, BF10 < 1 indicate larger evidence for the null hypothesis. For example, a BF10 = 3 means that the alternative hypothesis is three times more likely than the null hypothesis, whereas the reciprocal BF10= .33 means that the null hypothesis is three times more likely than the alternative. Following the guidelines by Jarosz and Wiley (2014), we consider values of BF10 = 1-3 (.1-.33) as anecdotal, BF10 = 3-10 (.33-.10) as moderate, BF10 = 10-30 (.10-.03) as strong, BF10 = 30-100 (.03-.01) as very strong and BF10 > 100 (< .01) as decisive evidence for the alternative hypothesis and the reciprocal values in parentheses as respective evidence for the null hypothesis.

In alignment with the approach by Nussbaum et al. (2024), we recoded responses in the PROMS from 0 to 1 in 0.25 steps starting with the “definitely” correct option down two the “definitely” incorrect option (thus, “don’t 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. For statistical analyses, we used the averaged performance across trials for each subtest.

## Transparency and openness

We specified how we determined our sample size, all data exclusions, all manipulations, and all measures in the associated preregistration (<https://doi.org/10.17605/OSF.IO/76PV5>). Preprocessed data, analysis scripts and supplemental materials can be found in the associated OSF repository (ToDo). For stimulus examples, refer to Nussbaum et al. (2024).

## Results

### Demography, musicality, and personality of participants

Singers and instrumentalists did not differ significantly in the socioeconomic status assessed via educational level, X2 (2, N = 88) = 1.06, *p* = .588; highest academic degree, X2 (7, N = 88) = 9.06, p = .249, and household income, X2 (4, N = 88) = 5.23, p = .264 (for more details see Table S2 on OSF). Further, the groups did not differ in age or positive and negative affect (assessed with the PANAS) and were comparable regarding Big Five personality traits and autistic traits. In the Gold-MSI, singers and instrumentalists scored comparatively on the general musicality score, but there were differences on two subfactors: instrumentalists scored higher on the subfactor Formal Education, while singers scored higher on Singing. In the PROMS, both groups performed comparably in all four subtests. Participant characteristics assessed via self-report and music performance in the PROMS are summarized in **Table 1**.

**Table 1**

*Characteristics of participants - Demography, personality, and musicality*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Singers** | **Instrumentalists** |  |  |  |  |  |
|  | M (SD) | M (SD) | t | dfa | p | Cohens d |  |
| Age | 27.02 (8.2) | 28.51 (10.6) | -0.73 | 78.93 | .465 | -0.17 [-0.61, 0.28] |  |
|  |  |  |  |  |  |  |  |
| *PANAS* |  |  |  |  |  |  |  |
| positive Affect | 3.00 (0.68) | 3.11 (0.57) | -0.77 | 84.78 | .446 | -0.17 [-0.59, 0.26] |  |
| negative Affect | 1.53 (0.47) | 1.40 (0.35) | 1.49 | 80.61 | .141 | 0.33 [-0.11, 0.77] |  |
|  |  |  |  |  |  |  |  |
| *Big Five* |  |  |  |  |  |  |  |
| Openness | 4.04 (0.55) | 3.99 (0.51) | 0.46 | 85.96 | .647 | 0.10 [-0.32, 0.52] |  |
| Conscientiousness | 3.47 (0.69) | 3.76 (0.70) | -1.91 | 85.62 | .060 | -0.41 [-0.84, 0.02] |  |
| Extraversion | 3.21 (0.70) | 3.00 (0.73) | 1.44 | 85.3 | .155 | 0.31 [-0.12, 0.74] |  |
| Agreeableness | 3.81 (0.57) | 4.01 (0.60) | -1.61 | 85.29 | .112 | -0.35 [-0.77, 0.08] |  |
| Neuroticism | 2.74 (0.77) | 2.61 (0.78) | 0.80 | 85.75 | .426 | 0.17 [-0.25, 0.60] |  |
|  |  |  |  |  |  |  |  |
| *AQ* |  |  |  |  |  |  |  |
| Total | 18.2 (6.15) | 19.28 (8.55) | -0.68 | 76.04 | .500 | -0.16 [-0.60, 0.30] |  |
| Attention to Detail | 5.4 (2.33) | 5.63 (2.53) | -0.44 | 84.64 | .662 | -0.10 [-0.52, 0.33] |  |
| Social | 12.8 (5.37) | 13.65 (7.53) | -0.61 | 75.65 | .545 | -0.14 [-0.59, 0.31] |  |
| Social Skills | 2.4 (1.94) | 3.09 (2.95) | -1.30 | 72.01 | .200 | -0.31 [-0.77, 0.16] |  |
| Communication | 2.53 (1.94) | 2.44 (2.3) | 0.20 | 82.02 | .842 | 0.04 [-0.39, 0.48] |  |
| Imagination | 2.51 (1.75) | 2.81 (1.88) | -0.78 | 84.86 | .437 | -0.17 [-0.60, 0.26] |  |
| Attention Switching | 5.36 (1.91) | 5.30 (2.23) | 0.12 | 82.69 | .904 | 0.03 [-0.40, 0.46] |  |
|  |  |  |  |  |  |  |  |
| *Gold-MSI* |  |  |  |  |  |  |  |
| General ME | 4.78 (0.85) | 4.75 (0.80) | 0.19 | 85.99 | .866 | 0.04 [-0.39, 0.46] |  |
| Active Engagement | 3.83 (0.82) | 4.21 (1.13) | -1.79 | 76.79 | .078 | -0.41 [-0.86, 0.05] |  |
| Formal Education | **4.39 (1.14)** | **4.95 (0.62)** | -2.85 | 68.31 | .006 | -0.69 [-1.18, -0.20] | \*\* |
| Emotion | 5.50 (0.81) | 5.60 (0.76) | -0.60 | 85.99 | .549 | -0.13 [-0.55, 0.29] |  |
| Singing | **4.98 (0.97)** | **4.19 (1.27)** | 3.25 | 78.56 | .002 | 0.73 [0.27, 1.19] | \*\* |
| Perception | 5.73 (0.82) | 5.77 (1.03) | -0.22 | 80.11 | .825 | -0.05 [-0.49, 0.39] |  |
|  |  |  |  |  |  |  |  |
| *PROMS* |  |  |  |  |  |  |  |
| Pitch | 0.23 (0.08) | 0.24 (0.06) | -0.30 | 82.92 | .766 | -0.07 [-0.50, 0.37] |  |
| Melody | 0.17 (0.10) | 0.14 (0.10) | 1.29 | 85.08 | .199 | 0.28 [-0.15, 0.71] |  |
| Timbre | 0.29 (0.08) | 0.3 (0.09) | -0.59 | 86.00 | .556 | -0.13 [-0.55, 0.30] |  |
| Rhythm | 0.31 (0.09) | 0.30 (0.08) | -0.56 | 85.13 | .577 | -0.12 [-0.55, 0.30] |  |

*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 score were calculated based on Hoekstra et al. (2008) and Baron-Cohen et al. (2001).*

*a Note that original degrees of freedom were 86 but were corrected due to unequal variance.*

### Emotion classification performance

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 (singers and instrumentalists) as a between subject factor (see **Table 2**).

The results revealed main effects of **Emotion** and **Morph Type**, which were qualified by an interaction. Crucially, however, we found no main effects or interactions involving **Group** (see **Figure 2**), which also confirmed in a Bayesian ANOVA (see OSF). Planned Bayesian analysis revealed moderate evidence for the null effect of group for overall performance (p = .542, BF10 = 0.265), as well as for Full (p = .392, BF10 = 0.310), F0 (p = .935, BF10 = 0.226), and Timbre morphs (p = .555, BF10 = 0.262) separately. Thus, we found evidence consistent with our hypotheses H1 and H2.

**Table 2**

*Results of the 4 × 3 × 2 mixed-effects ANOVA on the mean proportion of correct responses*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **df1** | **df2** | **F** | **p** | **Ωp2 [95%-CI]** | ***ε*HF** |
| Group | 1 | 86 | 0.38 | .542 | .00 [.00 .01] |  |
| Emotion | 3 | 258 | 72.43 | <.001 | .45 [.36 .53] |  |
| Morph Type | 2 | 172 | 768.93 | <.001 | .90 [.87 .93] | .741 |
| Group x Emotion | 3 | 258 | 2.14 | .095 | .01 [.00 .04] |  |
| Group x Morph Type | 2 | 172 | 0.36 | .635 | .00 [.00 .01] |  |
| Emotion x Morph Type | 6 | 516 | 22.78 | <.001 | .20 [.14 .25] | 827 |
| Group x Emotion x Morph Type | 6 | 516 | 1.33 | .249 | .00 [.00 .01] |  |

**Figure 2**

Mean proportion of correct responses per Morph Type separately for singers and instrumentalists



Note. Whiskers represent 95% confidence intervals. Violin plots represent variation of individual participants. The dotted line represents guessing rate at .25.

Follow-up analysis of the Morph Type effect revealed that performance was best in the Full condition (*M* = 0.75 ± 0.01 SEM), followed by the F0 (*M* = 0.62 ± 0.01) and then the Timbre condition (*M* = 0.41 ± 0.01); Full vs. F0: |*t*(87)| = 23.28, *p* < .001, *d* = 2.50 [2.07, 2.92], F0 vs Timbre: |*t*(87)| = 21.44, *p* < .001, *d* = 2.30 [1.90, 2.70], Full vs Timbre: |*t*(87)| = 33.94, *p* < .001, *d* = 3.64 [3.06, 4.21]). This Morph Type main effect was also found for all emotions separately (all *F*s(2, 174) > 102.44, *p* < .001), although it differed slightly between emotions, as suggested by the interaction (see **Figure 3,** for all post-hoc tests, refer to OSF).

To address our specific interest in the relative importance of F0 and Timbre for the different emotions, we calculated the performance differenceF0-Tbr for each emotion separately. Performance difference was largest for Happiness (M = 0.34 ± 0.02 SEM), followed by Fear (M = 0.21 ± 0.02), Sadness (M = 0.18 ± 0.02), and Pleasure (M = 0.10 ± 0.02; all pairwise comparisons |*t*s(77)| ≥ 2.57, ps≤ .012, ds ≥ 0.28 [0.06 0.49], except for Fear vs. Sadness (|t(87)| = 1.13, p = .261). These effects of Morph Type and Emotion therefore present a full replication of the patterns reported in Nussbaum et al. (2024). For confusion matrices and supplemental analyses, please refer to OSF.

**Figure 3**

Mean proportion of correct responses per Emotion and Morph Type



Note. Whiskers represent 95%-confidence intervals. Grey dots represent individual participants’ data. The dotted line represents guessing rate at .25.

# Part II: Correlational analyses

## Hypotheses

In Part II, we focused on the correlations between auditory sensitivity and vocal emotion recognition. We aimed to replicate the patterns found in Nussbaum et al. (2024), and therefore formulated the following hypotheses:

*Correlations with the PROMS*

**H5:** Averaged vocal emotion recognition (VER) performance is correlated with averaged music perception performance.

**H6:** Full-VER and F0-VER are correlated with melody perception in music

*Correlations with the GOLD-MSI:*

**H7:** Averaged-VER and Full-VER are correlated with the General-ME.

**H8:** Averaged-VER and Full-VER are correlated with the Perception Subscale

**H9:** Averaged-VER and Full-VER are correlated with self-rated singing abilities.

## Data analysis

We calculated Spearman correlations between vocal emotion perception performance and both the PROMS music perception performance and the Gold-MSI self-rated musicality. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). Following our pre-registered plan, all correlations were controlled for formal musical education.

## Results

Replicating our previous findings, we obtained a strong correlation between vocal emotion recognition and music perception performance, as measures with the PROMS (Table 3). Further, we replicated the specific link between vocal emotion perception and the melody subtest. Although we did not have a specific hypothesis, we again found a link with Rhythm perception. There were no links between performance in the timbre morph condition and the timbre subtest. Overall, this pattern of correlation provides an almost complete replication of the patterns observed the previous sample (Nussbaum et al.) 2024, Table 2, page 12. Thus, the link between auditory sensitivity and vocal emotion perception seems to be highly comparable across professional musicians, non-musicians and amateurs.

**Table 3**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PROMSAvg | Pitch | Melody | Timbre | Rhythm |
| VERAvg | **0.38 (0.003)** | 0.15 (0.204) | **0.27 (0.023)** | 0.22 (0.066) | **0.36 (0.003)** |
| Full-Morphs | **0.35 (0.005)** | 0.14 (0.212) | **0.28 (0.023)** | 0.23 (0.058) | **0.32 (0.008)** |
| F0-Morphs | **0.39 (0.003)** | 0.15 (0.204) | **0.34 (0.006)** | 0.24 (0.053) | **0.32 (0.008)** |
| Timbre-Morphs | 0.18 (0.124) | 0.08 (0.503) | 0.05 (0.673) | 0.08 (0.499) | 0.22 (0.062) |

*Note. VER = Vocal Emotion Recognition performance.* *p-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995)*

In contrast, we found no links between vocal emotion recognition performance and self-rated musicality, as measured by the Gold-MSI (all correlations ≤ 0.21, ps ≥ .051, both with and without correction for formal musical education, details on OSF). Thus, for amateurs, we could not replicate the link with self-rated musical sophistication, perception and singing abilities which we observed in our previous sample of professional musicians and non-musicians. Therefore, we found evidence for our hypotheses H5 and H6, but not for the hypotheses H7 – H9.

# Part III: Comparison of professionals, amateurs and non-musicians

In the third part, we compared amateur musicians to both professional musicians as well as non-musicians. In principle, we predicted that amateurs and professional musicians would be comparable regarding vocal emotion perception. However, as the evidence reviewed above showed that amateurs can differ from professionals in cognitive abilities which could be linked to emotional sensitivity, we also considered the option that amateurs could be more proficient at making emotional inferences than professionals. Compared to our group of non-musicians, we assumed that amateurs would outperform them when emotion were expressed via full emotion cues and F0 cues only, but not timbre, because this is exactly the pattern we observed for professional musicians in Nussbaum et al. (2024).

## Hypotheses

**H3:** Amateur musicians outperform non-musicians in vocal emotion perception, in the Full and in the F0 condition.

**H4:** Amateurs perform equal or better to professional musicians in the Full and the F0 condition.

## Method

For this analysis, we collapsed all participants from Part I into the group of amateur musicians and compared it to the groups of professional musicians and non-musicians reported in Nussbaum et al. (2024). Note that we added one participant to the professional group, because he held a master’s degree in music (see Part I), so numbers slightly diverge from the original publication. All professional musicians reported to have a music-related academic degree or a non-academic music qualification. Non-musicians were required not to play an instrument or engage in any other musical activities. For a more detailed description, please refer to Nussbaum et al. (2024).

In total, we analyzed data from 40 professional musicians (20 male, 20 female, aged 20 to 42 years [*M* = 29.6; *SD* = 5.58]), 38 non-musicians (18 male, 20 female, aged 19 to 48 years [*M* = 30.5; *SD* = 6.54]) and 88 amateurs (40 male, 46 female, 2 diverse, aged 18 to 54 years [*M* = 27.8; *SD* = 9.44].

The stimulus material, design and data analysis were identical to Part I. We focused our analysis on the comparison of amateurs with the other two groups, because the comparison of professional musicians and non-musicians is reported in Nussbaum et al. (2024).

## Results

### Demography, musicality, and personality of participants

Professionals, amateurs and non-musicians did not differ in the socioeconomic status assessed via educational level (χ²(6, N = 166) = 11.11, p = .085) and highest academic degree (χ²(16, N = 166) = 24.04, p = .089). However, there were differences regarding household income (χ²(8, N = 166) = 20.19, p = .010, Cramer’s V = .25), with amateurs reporting higher household income than professionals and non-musicians.

Participant characteristics are summarized in **Table 4.** For a full report of statistical details, please refer to OSF. The groups were comparable in age as well as in positive and negative affect (assessed with the PANAS). For the Big Five, analyses of variance revealed group differences for extraversion, with slightly higher levels in professionals than in amateurs. Regarding autistic traits, the three groups did not differ in their overall score, but there were differences on the several subscales. In the Gold-MSI, professional musicians scored significantly higher than amateurs, except Emotion, which in turn scored higher than non-musicians. This is a pattern (professionals > amateurs > non-musicians) one would expect for self-rated musicality. In the PROMS, professionals outperformed amateurs in the Pitch and Melody subtest, whereas there were no differences in the Timbre and Rhythm subtests. Amateurs performed better than non-musicians in the Pitch, Melody and Rhythm subtest but not in the Timbre subtest. Thus, a clear pattern of professionals > amateurs > non-musicians was only found for melody and pitch.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Professionals** |  | **Amateurs** |  | **Non-Musicians** |
|  | M (SD) |  | M (SD) |  | M (SD) |
| *PANAS* |  |  |  |  |  |
| positive Affect | 3.32 (0.65) |  | 3.05 (0.63) |  | 3.1 (0.67) |
| negative Affect | 1.69 (0.48) |  | 1.47 (0.42) |  | 1.49 (0.69) |
|  |  |  |  |  |  |
| *Big Five* |  |  |  |  |  |
| Openness | 4.12 (0.50) |  | 4.02 [0.53) |  | 3.81 (0.80) |
| Conscientiousness | 3.49 (0.71) |  | 3.61 (0.70) |  | 3.76 (0.72) |
| Extraversion | **3.48 (0.66)** | > | **3.11 (0.72)** |  | 3.38 (0.79) |
| Agreeableness | 3.92 (0.57) |  | 3.91 (0.59) |  | 3.75 (0.66) |
| Neuroticism | 2.95 (0.65) |  | 2.69 (0.77) |  | 2.58 (0.82) |
|  |  |  |  |  |  |
| *AQ* |  |  |  |  |  |
| Total | 15.7 (4.98) |  | 18.73 (7.40) |  | 17.58 (6.41) |
| Attention to Detail | 5.43 (2.04) |  | **5.51 (2.42)** | > | **4.32 (2.01)** |
| Social | **10.28 (4.70)** | < | **13.22 (6.49)** |  | 13.26 (6.51) |
| Social Skills | **1.48 (1.68)** | < | **2.74 (2.49)** |  | 2.61 (2.63) |
| Communication | 1.85 (1.61) |  | 2.49 (2.12) |  | 2.39 (1.73) |
| Imagination | 2.18 (1.52) |  | 2.66 (1.8) |  | 2.87 (1.95) |
| Attention Switching | 4.78 (1.91) |  | 5.33 (2.06) |  | 5.39 (1.92) |
|  |  |  |  |  |  |
| *Gold-MSI* |  |  |  |  |  |
| General ME | **5.68 (0.50)** | > | **4.76 (0.82)** | > | **2.74 (1.07)** |
| Active Engagement | **4.94 (0.81)** | > | **4.02 (1.00)** | > | **2.95 (1.19)** |
| Formal Education | **5.95 (0.56)** | **>** | **4.66 (0.96)** | > | **1.71 (0.68)** |
| Emotion | 5.88 (0.73) |  | **5.55 (0.78)** | > | **4.95 (1.32)** |
| Singing | **5.34 (0.83)** | **>** | **4.59 (1.19)** | > | **2.84 (1.26)** |
| Perception | **6.31 (0.51)** | > | **5.75 (0.92)** | > | **4.22 (1.49)** |
|  |  |  |  |  |  |
| *PROMS* |  |  |  |  |  |
| Pitch | **0.27 (0.06)** | > | **0.24 (0.07)** | > | **0.18 (0.06)** |
| Melody | **0.23 (0.08)** | > | **0.16 (0.10)** | > | **0.07 (0.08)** |
| Timbre | 0.32 (0.08) |  | 0.29 (0.08) |  | 0.26 (0.09) |
| Rhythm | 0.33 (0.08) |  | **0.32 (0.09)** | > | **0.27 (0.08)** |

**Table 4**

*Characteristics of participants - Demography, personality, and musicality*

*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 score were calculated based on Hoekstra et al. (2008) and Baron-Cohen et al. (2001). Comparison signs (“>” or “<”) indicate significant differences. For a full report of statistical details, please refer to OSF.*

### Emotion classification performance

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 (professionals, amateurs, and non-musicians) as a between subject-factor (see **Table 5**).

The results revealed the main effects of **Emotion** and **Morph Type**, which were qualified by interaction. Crucially, however, we found no main effect involving **Group** and only a trend for an interaction with **Morph Type**.

**Table 5**

*Results of the 4 × 3 × 3 mixed-effects ANOVA on the mean proportion of correct responses*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **df1** | **df2** | **F** | **p** | **Ωp2 [95%-CI]** | ***ε*HF** |
| Group | 2 | 163 | 1.96 | .144 | .01 [.00 .06] |  |
| Emotion | 3 | 489 | 130.24 | <.001 | .44 [.38 .49] |  |
| Morph Type | 2 | 326 | 1357.80 | <.001 | .89 [.87 .91] | .829 |
| Group x Emotion | 6 | 489 | 1.17 | .322 | .00 [.00 .00] |  |
| Group x Morph Type | 4 | 326 | 2.14 | .089 | .01 [.00 .04] | .829 |
| Emotion x Morph Type | 6 | 978 | 40.95 | <.001 | .20 [.15 .24] | .875 |
| Group x Emotion x Morph Type | 12 | 978 | 0.74 | .688 | .00 [.00 .00] | .865 |

Planned comparisons between **professionals and amateurs** revealed moderate evidence for the null effect of overall performance (p = .473, BF10 = 0.238), as well as for Full (p = .322, BF10 = 0.286), F0 (p = .435, BF10 = 245), and Timbre morphs (p = .840, BF10 = 0.205) separately. Planned comparisons between amateurs and non-musicians revealed inconclusive evidence for overall performance (p = .107, BF10 = 0.546), as well as for Full (p = .044, BF10 = 1.009), and F0 (p = .105, BF10 = 0.576) separately. For Timbre morphs, there was moderate evidence for the null effect (p = .975, BF10 = 0.205). Thus, we found evidence consistent with our hypotheses H4, but inconclusive evidence regarding H3.

**Figure 4**

Mean proportion of correct responses per Morph Type separately for professionals, amateurs, and non-musicians



# Discussion

In the present study, we shed new light on the link between musicality and vocal emotion perception, by focusing on different subgroups of musicians. In line with our hypotheses, emotion recognition was found to be comparable between both singers and instrumentalists, as well as between amateur musicians and professionals. We further replicated the consistent link between sensitivity towards musical patterns and vocal emotion perception, although this was reflected in objective performance measures only, but not in self-rated musicality. In total, these results suggest that the link between musicality and vocal emotion perception is driven by individual differences in auditory sensitivity, which is not tied to a particular type or amount of musical activity. In what follows, we will discuss these findings in more detail.

## Singers vs. instrumentalists

For the present study, we recruited a well-powered sample of singers and instrumentalists, which were comparable with regard to personality, socio-economic background, and objective music perception performance. They only differed in two aspects of self-rated musicality: unsurprisingly, singers scored higher on self-rated singing abilities and instrumentalists scored higher on self-rated formal education. The latter is because mastering an instrument to the point where one can play in an ensemble arguably takes more formal training than being able to sing in a choir.

In line with our prediction, vocal emotion recognition performance of singers and instrumentalists did not differ in any condition, with moderate evidence quantified via Baysian analysis. This finding does not represent a mere absence of effects. In fact, we replicated the strong effect of our voice morphing condition on vocal emotion recognition, with F0 contour being relatively more informative than timbre, across all emotions, while performance was still best in the Full condition, where both F0 and timbre information informative. The fact that this pattern was found to be highly comparable for singers and instrumentalists suggests similar profiles of vocal emotional processing.

Although this is precisely what we predicted, we want to address three observations from the literature which may seem incompatible with our findings: First, production and perception is tightly linked in emotional communication (Frühholz & Schweinberger, 2021; Schirmer et al., 2025), so assuming that singers would outperform instrumentalists seems plausible on first sight. Note that our findings are not in conflict with the principle notion that there is a strong link between perception and production in auditory communication. They simply challenge the assumption that this link is different in singers and instrumentalists, similar to previous studies (Nikjeh et al., 2009). Second, and related to the previous point, several studies found correlations between singing abilities and emotion recognition (Correia et al., 2022; Greenspon & Montanaro, 2023; Nussbaum et al., 2024). However, all these studies included both singers and instrumentalists. So again, we do not argue against such a link per se, but we lack evidence that this is more pronounced in specific groups of musicians. In fact, we had predicted a correlation between self-rated singing abilities and emotion recognition as well but failed to find it in the present study. We attribute this to potentially decreased variance in our group of amateurs with regard to self-rated musicality in general, because we did find any correlations with these measures. Thus, we still agree with the notion from the literature that singing abilities and emotion perception are linked. Third, our findings diverge from Thompson et al. (2004), who found that singing lessons may even interfere with vocal emotional processing, but this study had several methodological limitations including a very small sample size. With the present, substantially powered design, we do not see any evidence for a disruptive effect of active singing. This aligns with the broader literature on protective effects of singing on cognitive and socio-emotional functioning (Moisseinen et al., 2024; Tragantzopoulou & Giannouli, 2025).

## Amateurs compared to professional musicians and non-musicians

In line with our prediction, we found evidence that emotion perception performance does not differ between amateur musicians and non-musicians, further strengthening the notion that the amount of music training is not a major influence on vocal emotion perception (Schellenberg & Lima, 2024). However, we also hypothesized that amateurs would outperform non-musicians, because professionals outperformed them in our previous study (Nussbaum et al., 2024). This comparison yielded inconclusive evidence. We speculate that the amateur sample was more heterogenous than the professional one, and as a result, our design lacked statistical power to detect potentially very small differences between amateurs and non-musicians, despite the substantial sample size. This may be resolved in a follow-up study.

In addition to the direct group comparison, we performed a correlational analysis on our sample of amateurs (Part II) which was in parallel to the one we had previously performed on professionals and non-musicians (cf. Table 2 and Table 3 in (Nussbaum et al., 2024)). Importantly, we found a highly similar pattern of correlations between vocal emotion perception and music perception abilities, especially for melody and rhythm. We therefore conclude that individual differences in music perception abilities play an important role, irrespective of the assignment to any (non)-musical group. This is fully in line with the current literature, which consistently emphasizes the role of acoustic sensitivity and argues against a causal effect of training or musical activity on vocal emotion perception (Neves et al., 2025; Schellenberg & Lima, 2024).

## Constraints on generality and future directions

To the best of our knowledge, this study is the first substantially powered comparison between different subgroups of musicians with respect to vocal emotion perception and therefore closes an open gap in the literature. Nevertheless, the present study has several limitations which deserve consideration for future research efforts. All musicians were socialized in Western music culture and fluent German speakers, therefore findings may not generalize to other music and language backgrounds (Morrison & Demorest, 2009). Note, however, that links between music perception performance and emotion recognition irrespective of formal musical training have been observed in a Portuguese sample as well (Correia et al., 2022).

What posed a particular challenge in the present study was the recruitment of mutually exclusive groups of musicians. While the distinction between singers and instrumentalists may seem trivial at first glance, it is actually rare that musicians engage in one of these activities only. Thus, for practical reasons, we recruited participants who self-assigned clearly to one predominant form and made sure that there current activity is either one or the other. In a similar vein, the distinction between amateurs and professionals is not fully straight-forward and may represent a continuum rather than clear-cut categories. Some individuals pursue music as a profession, but without a formal music degree, others vice versa (Zendel & Alexander, 2020). Some people transition from amateurs to a professional level, even later at life (Taylor & Hallam, 2011). Thus, while the argument still holds that future research should incorporate the heterogeneity of musicians, it may be more adequate to consider variability on a spectrum rather than distinct groups. We aimed to incorporate this approach in our correlational analysis of Part II.

On a practical note, we must acknowledge that the technical randomization error limits comparability of this data collection encompassing amateur singers and instrumentalists with the previous one encompassing professionals and non-musicians. While in the first study stimuli were drawn only once, as intended, the present code allowed full randomization with duplication and omissions of stimuli. While undoubtedly unfortunate, we are nevertheless confident that this error has not affected our results to a large degree. First, the classification patterns for different Morph Types and Emotions present a full replication of the previous study (cf Figure 3) and we observed very similar correlations between vocal emotion recognition and music perception performance (cf. Table 3). Second, while this issue could have arguably decreased signal-to-noise ratio, we do not see a risk that it could have introduced a specific bias. Thus, we still consider both studies sufficiently comparable.

While the present behavioral data did not reveal any differences between musical subgroups, future research may unravel more fine-grained patterns in the brain. Several studies have reported brain differences in auditory and motor processing between amateurs and professionals (Kleber et al., 2010; Lotze et al., 2003; Oechslin et al., 2013; Papadaki et al., 2023) and singers and instrumentalists (Halwani et al., 2011), but specific insights into vocal emotional processing are elusive. With the present stimulus material, we observed distinct electrophysiological responses in musicians and non-musicians (Lehnen et al., 2025; Nussbaum et al., 2023), but a meaningful analysis of musical subgroups would have required a much bigger sample.

Finally, with respect to the individual dynamics of the action-perception link in vocal communication, it may be worthwhile to focus on another group of vocal experts: professional voice actors or imitators. A recent study found that voice actors show enhanced sensitivity for linguistic voice prosody and enhanced neural tracking for voice pitch compared to non-actors (Kachlicka & Tierney, 2024). Vocal acting usually involves deliberate exaggeration of vocal expressions and may therefore require fine-grained explicit representation of vocal emotions (Berry & Brown, 2019). While there is research comparing actors and non-actors regarding expression of emotions (Jürgens et al., 2015), it is not yet understood how this is mirrored in vocal emotion perception performance and how this group may differ from individuals with singing expertise.

# Summary and Conclusion

In the present study, we investigated how emotion perception differs in musical subgroups and compared singers and instrumentalists as well as amateurs and professionals. In line with our prediction, we observed no differences between musica subgroups, suggesting that emotion perception abilities are not primarily tied to the type or amount of musical activity. Instead, we replicated the consistent link between vocal emotion perception and music perception abilities, especially for melodies. This adds a new perspective to the accumulating evidence that the link between musicality and vocal emotion perception is predominantly driven by individual differences in auditory sensitivity, rather than specific musical training.

# Acknowledgements

This study was conducted by J.D. in partial fulfilment of the requirements for a master’s thesis. 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. We are grateful to all participants of the study.

# Conflicts of Interests and Funding

The authors declare no conflicts of interests.

# Credit Author Statement

Christine Nussbaum – Conceptualization, Methodology, Software, Visualization, Formal analysis, Writing - Original Draft, Supervision

Jessica Dethloff - Data collection, Formal analysis, Visualization, Writing - Original Draft

Annett Schirmer – Methodology, Writing - Review & Editing, Supervision

Stefan R. Schweinberger – Conceptualization, Writing - Review & Editing, Supervision

# Supplementary material

Supplemental figures and tables, analysis scripts, and raw data can be found on the associated OSF repository (ToDo).

References

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., Busch, V., Lothwesen, K., Elvers, P., Fischinger, T., Schlemmer, K., & Frieler, K. (2019). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition & Emotion*, *33*(6), 1099–1118. https://doi.org/10.1080/02699931.2018.1541312

Baldé, A. M., Lima, C. F., & Schellenberg, E. G. (2025). Associations between musical expertise and auditory processing. *Journal of Experimental Psychology: Human Perception and Performance.* Advance online publication. https://doi.org/10.1037/xhp0001312

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

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

Berry, M., & Brown, S. (2019). Acting in action: Prosodic analysis of character portrayal during acting. *Journal of Experimental Psychology. General*, *148*(8), 1407–1425. https://doi.org/10.1037/xge0000624

Bonde, L. O., Juel, K., & Ekholm, O. (2018). Associations between music and health-related outcomes in adult non-musicians, amateur musicians and professional musicians—Results from a nationwide Danish study. *Nordic Journal of Music Therapy*, *27*(4), 262–282. https://doi.org/10.1080/08098131.2018.1439086

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

Christiner, M., & Reiterer, S. M. (2015). A Mozart is not a Pavarotti: Singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience*, *9*, 482. https://doi.org/10.3389/fnhum.2015.00482

Correia, A. I., Castro, S. L [Sao Luis], 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

Fisher, R. A., Hoult, A. R., & Tucker, W. S. (2020). A Comparison of Facial Muscle Activation for Vocalists and Instrumentalists. *Journal of Music Teacher Education*, *30*(1), 53–64. https://doi.org/10.1177/1057083720947412

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rösler, M., & Gontard, A. von (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., & Schweinberger, S. R. (2021). Nonverbal auditory communication - Evidence for integrated neural systems for voice signal production and perception. *Progress in Neurobiology*, *199*, 101948. https://doi.org/10.1016/j.pneurobio.2020.101948

Greenspon, E. B., & Montanaro, V. (2023). Singing ability is related to vocal emotion recognition: Evidence for shared sensorimotor processing across speech and music. *Attention, Perception & Psychophysics*, *85*(1), 234–243. https://doi.org/10.3758/s13414-022-02613-0

Hake, R., Kreutz, G., Frischen, U., Schlender, M., Rois-Merz, E., Meis, M., Wagener, K. C., & Siedenburg, K. (2024). A Survey on Hearing Health of Musicians in Professional and Amateur Orchestras. *Trends in Hearing*, *28*, 23312165241293762. https://doi.org/10.1177/23312165241293762

Halwani, G. F., Loui, P., Rüber, T., & Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. *Frontiers in Psychology*, *2*, 156. https://doi.org/10.3389/fpsyg.2011.00156

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. *J Autism Dev Disord*, *38*(8), 1555–1566. https://doi.org/10.1007/s10803-008-0538-x

Jarosz, A. F., & Wiley, J. (2014). What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving*, *7*(1). https://doi.org/10.7771/1932-6246.1167

Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of Acting Experience on Emotion Expression and Recognition in Voice: Non-Actors Provide Better Stimuli than Expected. *Journal of Nonverbal Behavior*, *39*(3), 195–214. https://doi.org/10.1007/s10919-015-0209-5

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol Bull*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kachlicka, M., & Tierney, A. (2024). Voice actors show enhanced neural tracking of pitch, prosody perception, and music perception. *Cortex*, *178*, 213–222. https://doi.org/10.1016/j.cortex.2024.06.016

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*.

Kleber, B [B.], Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The brain of opera singers: Experience-dependent changes in functional activation. *Cerebral Cortex*, *20*(5), 1144–1152. https://doi.org/10.1093/cercor/bhp177

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

Krishnan, S., Lima, C. F., Evans, S., Chen, S., Guldner, S., Yeff, H., Manly, T., & Scott, S. K. (2018). Beatboxers and Guitarists Engage Sensorimotor Regions Selectively When Listening to the Instruments They can Play. *Cerebral Cortex*, *28*(11), 4063–4079. https://doi.org/10.1093/cercor/bhy208

Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *J Pers Soc Psychol*, *111*(5), 686–705. https://doi.org/10.1037/pspi0000066

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

Lehnen, J. M., Schweinberger, S. R., & Nussbaum, C. (2025). Vocal Emotion Perception and Musicality-Insights from EEG Decoding. *Sensors (Basel, Switzerland)*, *25*(6). https://doi.org/10.3390/s25061669

Lotze, M., Scheler, G., Tan, H.‑R. M., Braun, C., & Birbaumer, N. (2003). The musician's brain: Functional imaging of amateurs and professionals during performance and imagery. *Neuroimage*, *20*(3), 1817–1829. https://doi.org/10.1016/j.neuroimage.2003.07.018

Loveday, C., Musgrave, G., & Gross, S.‑A. (2023). Predicting anxiety, depression, and wellbeing in professional and nonprofessional musicians. *Psychology of Music*, *51*(2), 508–522. https://doi.org/10.1177/03057356221096506

Maghiar, M. J., Lawrence, B. J., Mulders, W. H., Moyle, T. C., Livings, I., & Jayakody, D. M. P. (2023). Hearing loss and mental health issues in amateur and professional musicians. *Psychology of Music*, *51*(6), 1584–1597. https://doi.org/10.1177/03057356231155970

Martins, I., Lima, C. F., & Pinheiro, A. P. (2022). Enhanced salience of musical sounds in singers and instrumentalists. *Cogn Affect Behav Neurosci.* Advance online publication. 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

Mithen, S., Morley, I., Wray, A., Tallerman, M., & Gamble, C. (2006). The Singing Neanderthals: the Origins of Music, Language, Mind and Body by Steven Mithen. London: Weidenfeld & Nicholson, 2005. ISBN 0-297-64317-7 hardback £20 & US$25.2; ix+374 pp. *Cambridge Archaeological Journal*, *16*(1), 97–112. https://doi.org/10.1017/S0959774306000060

Moisseinen, N., Ahveninen, L., Martínez-Molina, N., Sairanen, V., Melkas, S., Kleber, B [Boris], Sihvonen, A. J., & Särkämö, T. (2024). Choir singing is associated with enhanced structural connectivity across the adult lifespan. *Human Brain Mapping*, *45*(7), e26705. https://doi.org/10.1002/hbm.26705

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

Neves, L., Martins, M., Correia, A. I., Castro, S. L [São Luís], Schellenberg, E. G., & Lima, C. F. (2025). Does music training improve emotion recognition and cognitive abilities? Longitudinal and correlational evidence from children. *Cognition*, *259*, 106102. https://doi.org/10.1016/j.cognition.2025.106102

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. *The Journal of the Acoustical Society of America*, *125*(1), 328–338. https://doi.org/10.1121/1.3021309

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2023). Electrophysiological Correlates of Vocal Emotional Processing in Musicians and Non-Musicians. *Brain Sciences*, *13*(11), 1563. https://doi.org/10.3390/brainsci13111563

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2024). Musicality - Tuned to the melody of vocal emotions. *Br J Psychol*, *115*(2), 206–225. https://doi.org/10.1111/bjop.12684

Nussbaum, C., & Schweinberger, S. R. (2021). Links Between Musicality and Vocal Emotion Perception. *Emotion Review*, *13*(3), 211–224. https://doi.org/10.1177/17540739211022803

Oechslin, M. S., van de Ville, D., Lazeyras, F., Hauert, C.‑A., & James, C. E. (2013). Degree of musical expertise modulates higher order brain functioning. *Cerebral Cortex*, *23*(9), 2213–2224. https://doi.org/10.1093/cercor/bhs206

Papadaki, E., Koustakas, T., Werner, A., Lindenberger, U., Kühn, S., & Wenger, E. (2023). Resting-state functional connectivity in an auditory network differs between aspiring professional and amateur musicians and correlates with performance. *Brain Structure & Function*, *228*(9), 2147–2163. https://doi.org/10.1007/s00429-023-02711-1

Rakei, A., & Bhattacharya, J. (2024). Professional status matters: Differences in flow proneness between professional and amateur contemporary musicians. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000674

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.* Advance online publication. https://doi.org/10.1027/1015-5759/a000481

Rogenmoser, L., Kernbach, J., Schlaug, G., & Gaser, C. (2018). Keeping brains young with making music. *Brain Structure & Function*, *223*(1), 297–305. https://doi.org/10.1007/s00429-017-1491-2

Rosenfeld, J. P., & Olson, J. M. (2021). Bayesian Data Analysis: A Fresh Approach to Power Issues and Null Hypothesis Interpretation. *Applied Psychophysiology and Biofeedback*, *46*(2), 135–140. https://doi.org/10.1007/s10484-020-09502-y

Schellenberg, E. G., & Lima, C. F. (2024). Music Training and Nonmusical Abilities. *Annu Rev Psychol*, *75*, 87–128. https://doi.org/10.1146/annurev-psych-032323-051354

Scherer, K. R. (2018). Acoustic Patterning of Emotion Vocalizations. In S. Frühholz, P. Belin, & K. R. Scherer (Eds.), *The Oxford Handbook of Voice Perception* (pp. 60–92). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.4

Schirmer, A., Croy, I., Liebal, K., & Schweinberger, S. R. (2025). Non-verbal effecting - animal research sheds light on human emotion communication. *Biological Reviews of the Cambridge Philosophical Society*, *100*(1), 245–257. https://doi.org/10.1111/brv.13140

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

Taylor, A., & Hallam, S. (2011). From leisure to work: amateur musicians taking up instrumental or vocal teaching as a second career. *Music Education Research*, *13*(3), 307–325. https://doi.org/10.1080/14613808.2011.603044

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

Tragantzopoulou, P., & Giannouli, V. (2025). A Song for the Mind: A Literature Review on Singing and Cognitive Health in Aging Populations. *Brain Sciences*, *15*(3). https://doi.org/10.3390/brainsci15030227

Vigl, J., Talamini, F., Strauss, H., & Zentner, M. (2024). Prosodic discrimination skills mediate the association between musical aptitude and vocal emotion recognition ability. *Scientific Reports*, *14*(1), 16462. https://doi.org/10.1038/s41598-024-66889-y

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

Zendel, B. R., & Alexander, E. J. (2020). Autodidacticism and Music: Do Self-Taught Musicians Exhibit the Same Auditory Processing Advantages as Formally Trained Musicians? *Frontiers in Neuroscience*, *14*, 752. https://doi.org/10.3389/fnins.2020.00752

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