Comparing vocal emotion perception in singers vs. instrumentalists and amateur vs. professional musicians

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

**Supplemental Tables and Figures**

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

# Supplemental Sample Information

**Table S1**

*List of reported instruments by amateurs (singers and instrumentalists)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Singers** | **Instrumentalists** | | | |
| Gesang *(singing)* / Chor *(choir)* | | 35 | Violine *(violin)* | 10 |
| + Klavier *(piano)* | | 3 | Posaune (*trombone*) | 6 |
| + Violine (*violin*) | | 2 | Cello (*cello*) | 5 |
| + Gitarre *(guitar)* | | 2 | Klarinette (*clarinet*) | 4 |
| + Bass (*bass*) | | 1 | Trompete (*trumpet*) | 4 |
| + Klavier *(piano)* and Cello *(cello)* | | 1 | Waldhorn (*horn*) | 2 |
| + Trompete (*trumpet*) | | 1 | Bratsche (*Viola*) | 2 |
|  | |  | Schlagzeug (*drums*) | 2 |
|  | |  | Bariton (*bariton*) | 1 |
|  | |  | Bass (*bass*) | 1 |
|  | |  | Fagott (bassoon) | 1 |
|  | |  | Gitarre *(guitar)* | 1 |
|  | |  | Klavier *(piano)* | 1 |
|  | |  | Querflöte (*flute*) | 1 |
|  | |  | Tuba (*tuba*) | 1 |
|  | |  | Saxophon (saxophone) | 1 |

**Table S2**

*Socioeconomic background of amateurs (singers and instrumentalists)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Income (in €)** |  |  | **Education** |  |  | **Degree** |  |  |
|  | **S** | **I** |  | **S** | **I** |  | **S** | **I** |
| <1750 | 5 | 4 | keine *(none)* | 0 | 0 | keine *(none)* | 0 | 1 |
| 1750-2500 | 6 | 7 | Schüler *(pupil)* | 0 | 1 | Schüler *(pupil)* | 0 | 0 |
| 2500-3500 | 11 | 10 | Hauptschule  *(secondary school)* | 0 | 0 | In Ausbildung  *(under training)* | 15 | 17 |
| 3500-5000 | 17 | 9 | Mittelschule  *(secondary school)* | 0 | 0 | Lehre  *(traineeship)* | 0 | 3 |
| >5000 | 6 | 13 | Fachschule  *(technical college)* | 1 | 1 | Fachschule  *(technical college)* | 1 | 0 |
|  |  |  | Abitur  *(A-levels)* | 44 | 41 | Meister  *(master craftsmen)* | 0 | 0 |
|  |  |  |  |  |  | Bachelor  *(Bachelor)* | 10 | 4 |
|  |  |  |  |  |  | Fachhochschulabschluss *(polytechnic degree)* | 1 | 2 |
|  |  |  |  |  |  | Master/Diplom *(Master/Diploma)* | 14 | 10 |
|  |  |  |  |  |  | Promotion *(PhD)* | 4 | 6 |
|  |  |  |  |  |  |  |  |  |
| χ2 = 5.23, df = 4, p = 0.264 | | | χ2 = 1.06, df = 2, p = 0.588 | | | χ2 = 9.06, df = 7, p = 0.249 | | |

*Note. This table presents the number of individuals belonging to different income, education, and degree categories. We tested group differences between musician (M) and non-musicians (C) using a Chi-square test and show the results in the last line of this table. Please note that the response options “Education” (i.e. the type of school) and “Degree” (i.e. the highest professional qualification) were tailored to the German educational system and are therefore difficult to translate Further, please note that “Fachschule” and “Abitur” are similar as they both enable a person to pursue a university degree (with a few more constrains for a “Fachschule” degree). We therefore consider the trend observed for the “Education” factor merely as an artefact of the response format. M = Musicians, C = Controls/Non-Musicians*

# Supplemental Stimulus Information

**Table S3**

*Summary of the acoustic characteristics of female voice morphs separately for each Emotion and Morph Type*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MType** | **F0 Mean** | **F0 SD** | **F0 Glide** | **FormDisp** | **HNR** |
| *Happiness* |  |  |  |  |  |  |
|  | Full | 348 | 98 | -112 | 993 | 19 |
|  | F0 | 348 | 98 | -112 | 1096 | 20 |
|  | Timbre | 247 | 25 | -37 | 981 | 19 |
| *Pleasure* |  |  |  |  |  |  |
|  | Full | 185 | 21 | -32 | 1131 | 19 |
|  | F0 | 185 | 21 | -32 | 1094 | 19 |
|  | Timbre | 247 | 25 | -37 | 1122 | 20 |
| *Fear* |  |  |  |  |  |  |
|  | Full | 288 | 30 | 28 | 1112 | 21 |
|  | F0 | 288 | 30 | 28 | 1093 | 21 |
|  | Timbre | 247 | 25 | -37 | 1120 | 21 |
| *Sadness* |  |  |  |  |  |  |
|  | Full | 219 | 19 | -39 | 1090 | 22 |
|  | F0 | 219 | 19 | -39 | 1097 | 21 |
|  | Timbre | 247 | 25 | -37 | 1085 | 22 |
| *Average* |  |  |  |  |  |  |
|  | Full | 247 | 25 | -39 | 1094 | 22 |

*Note. All acoustical parameters were adapted from (McAleer et al., 2014) and extracted using Praat software (Boersma, 2018) and the F0 contour information from the TANDEM-STRAIGHT object in Matlab (MATLAB, 2020). F0 Glide = F0End – F0Start; Formant Dispersion (FormDisp) = ratio between consecutive formant means (from F1 to F4, maximum formant frequency set to 5.5 kHz, window length 0.025 s); HNR (harmonics-to-noise ratio) was extracted with the cross-correlation method (mean value; time step = 0.01 s; min pitch = 75 Hz; silence threshold = 0.1, periods per window = 1.0).*

**Table S4**

*Summary of the acoustic characteristics of male voice morphs separately for each Emotion and Morph Type*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MType** | **F0 Mean** | **F0 SD** | **F0 Glide** | **FormDisp** | **HNR** |
| *Happiness* |  |  |  |  |  |  |
|  | Full | 259 | 89 | -74 | 999 | 17 |
|  | F0 | 259 | 89 | -74 | 1037 | 17 |
|  | Timbre | 158 | 21 | -43 | 985 | 15 |
| *Pleasure* |  |  |  |  |  |  |
|  | Full | 121 | 18 | -32 | 1064 | 14 |
|  | F0 | 121 | 18 | -32 | 1046 | 15 |
|  | Timbre | 158 | 21 | -43 | 1058 | 14 |
| *Fear* |  |  |  |  |  |  |
|  | Full | 191 | 23 | -19 | 1077 | 17 |
|  | F0 | 191 | 23 | -19 | 1046 | 17 |
|  | Timbre | 158 | 21 | -43 | 1074 | 17 |
| *Sadness* |  |  |  |  |  |  |
|  | Full | 122 | 14 | -47 | 1040 | 16 |
|  | F0 | 122 | 14 | -47 | 1049 | 16 |
|  | Timbre | 158 | 21 | -43 | 1033 | 16 |
| *Average* |  |  |  |  |  |  |
|  | Full | 158 | 21 | -43 | 1047 | 17 |

*Note. All acoustical parameters were adapted from (McAleer et al., 2014) and extracted using Praat software (Boersma, 2018) and the F0 contour information from the TANDEM-STRAIGHT object in Matlab (MATLAB, 2020). F0 Glide = F0End – F0Start; Formant Dispersion (FormDisp) = ratio between consecutive formant means (from F1 to F4, maximum formant frequency set to 5.5 kHz, window length 0.025 s); HNR (harmonics-to-noise ratio) was extracted with the cross-correlation method (mean value; time step = 0.01 s; min pitch = 75 Hz; silence threshold = 0.1, periods per window = 1.0).*

# Supplemental Design Information

**Table S5**

*Summary of response key mappings to emotions.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **„d“** | **„f“** | **„j“** | **„k“** |
| CB 1 | happiness | pleasure | sadness | fear |
| CB 2 | sadness | fear | happiness | pleasure |
| CB 3 | pleasure | happiness | fear | sadness |
| CB 4 | fear | sadness | pleasure | happiness |

*Note. Participants were instructed explicitly to press the keys „d“ and „f“ with their left index- and middle-finger and the keys „j“ and „k“ with their right index- and middle-finger. CB = counterbalancing condition.*

**Table S6**

*Participant assignment to the different response key mappings.*

|  |  |  |
| --- | --- | --- |
|  | **Singers** | **Instrumentalists** |
| CB 1 | 11 | 10 |
| CB 2 | 10 | 12 |
| CB 3 | 16 | 8 |
| CB 4 | 8 | 13 |

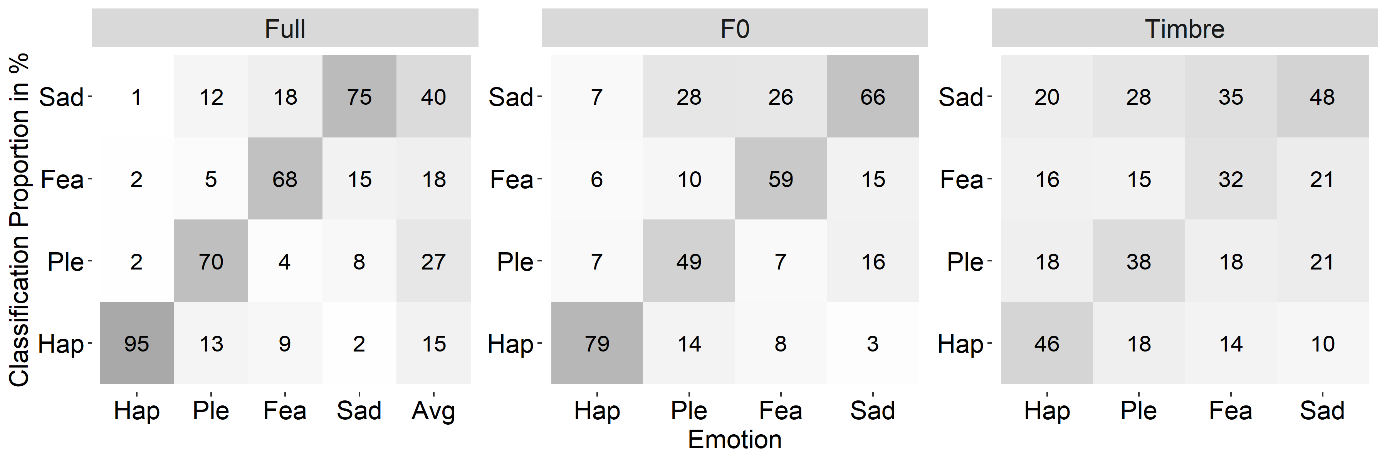
*Note. Participants were randomly assigned to key mappings. CB = counterbalancing condition.*

*ToDo ab hier*

# Supplemental Results

**Figure S1**

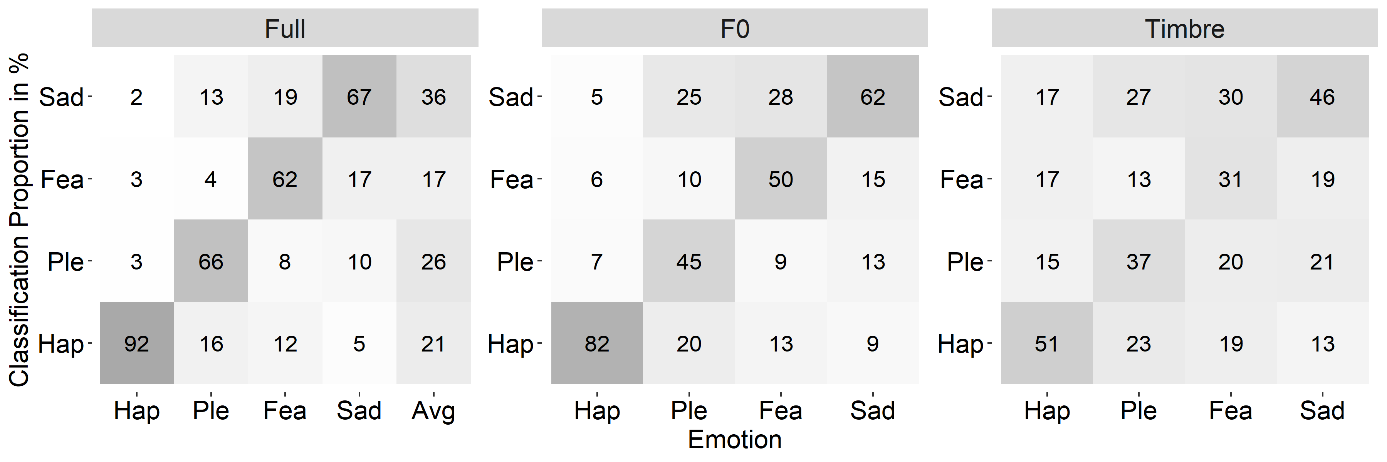
*Confusion data for each Emotion for the three Morph Types* *– Musicians only*

**

*Note. Numbers represent the proportion of classification responses per Emotion and Morph Type, averaged across musicians. Hap = happiness, Ple = pleasure, Fea = fear, Sad = sadness, Avg = average.*

**Figure S2**

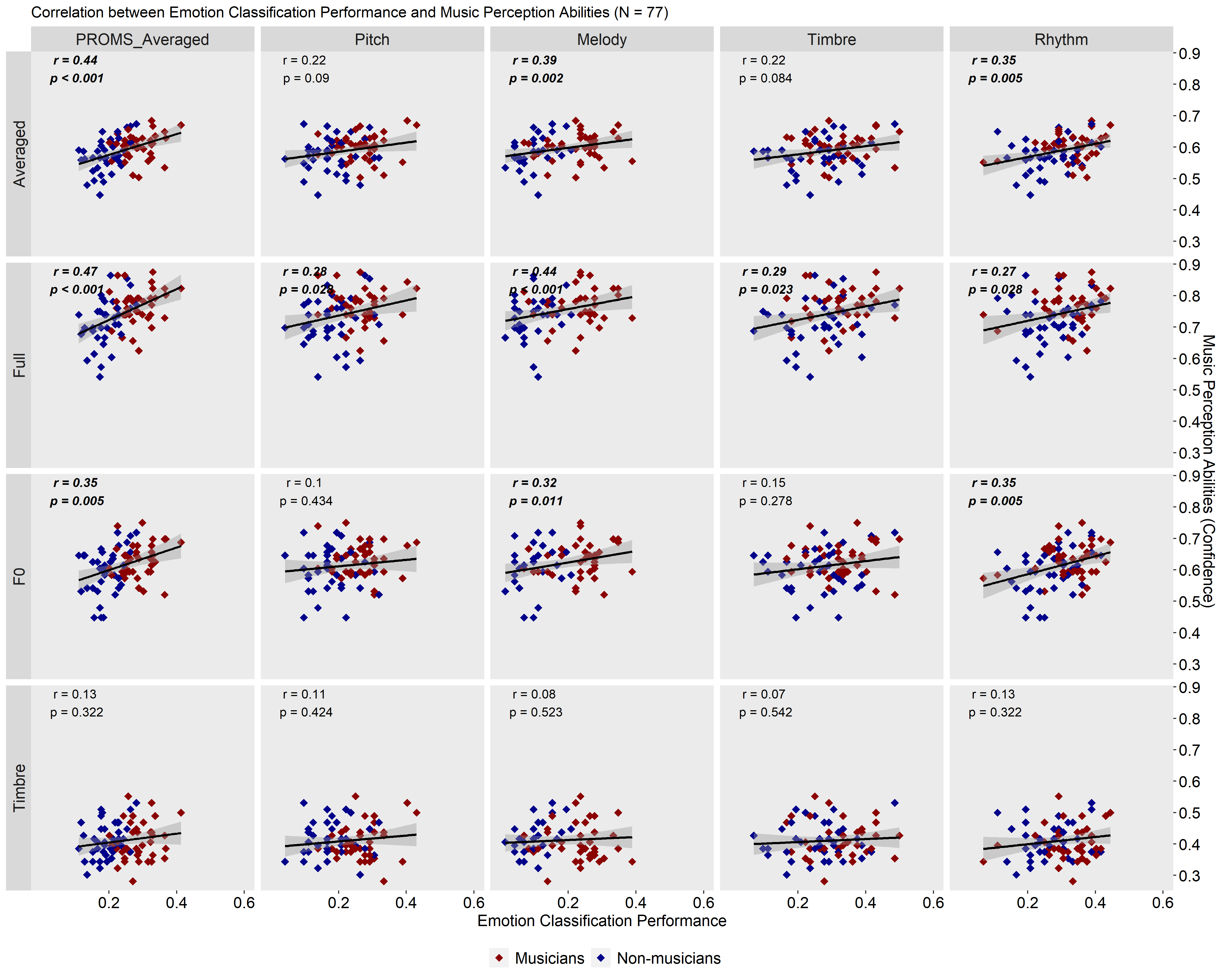
*Confusion data for each Emotion for the three Morph Types – Non-Musicians only*



*Note. Numbers represent the proportion of classification responses per Emotion and Morph Type, averaged across non-musicians. Hap = happiness, Ple = pleasure, Fea = fear, Sad = sadness, Avg = average.*

**Figure S3**

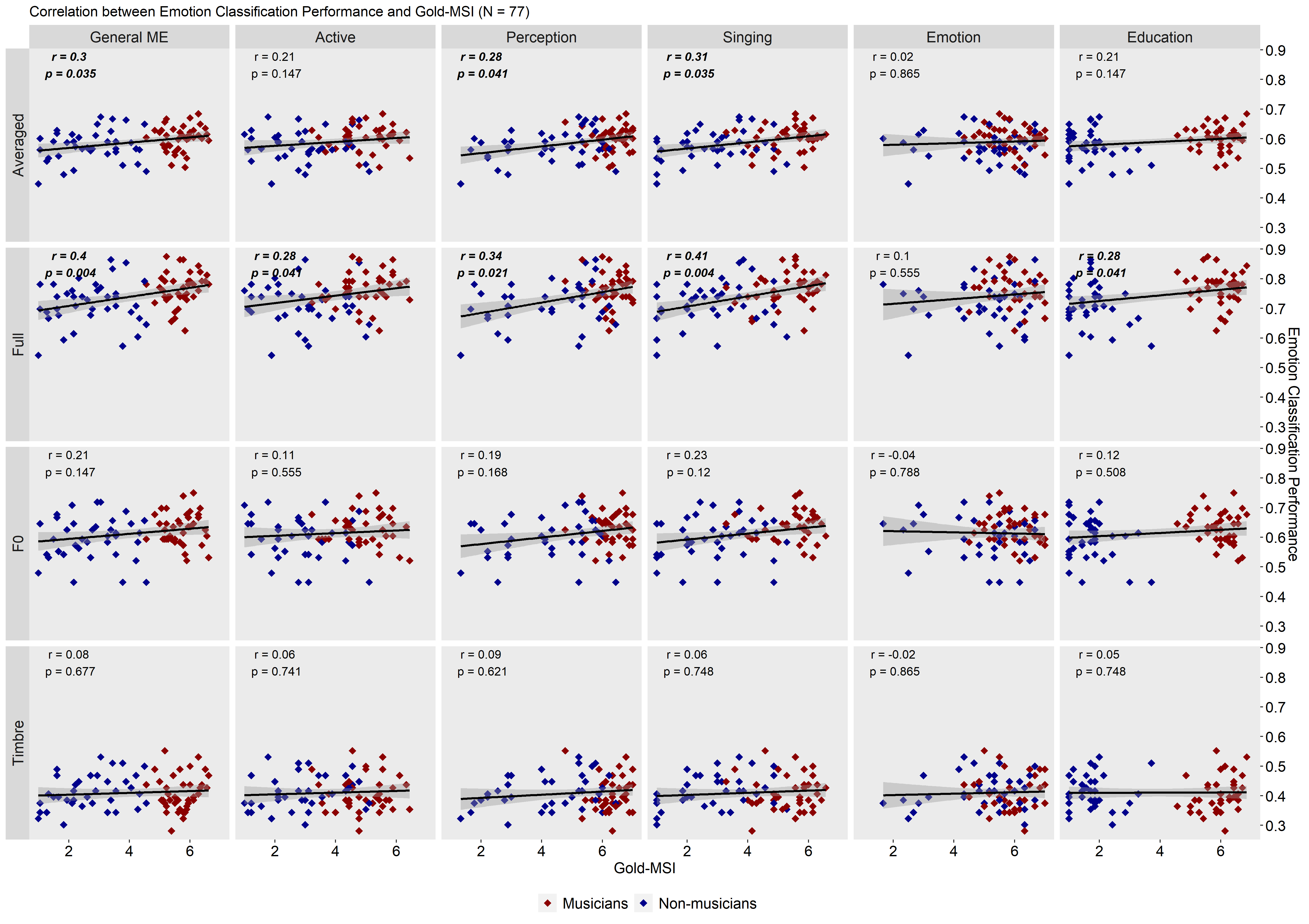
*Correlation between Emotion Classification Performance and Music Perception Abilities (PROMS)*

**

*Note. The x-axis shows the different subtests of the PROMS (Pitch, Melody, Timbre, and Rhythm) as well as the averaged performance across all subtests (PROMS\_Averaged). The y-axis shows the vocal emotion classification performance separately for each Morph Type (Full, F0 and Timbre) and averaged across Morph Types (Averaged). p-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995; false discovery rate set to 0.05, total number of tests = 20)*

**Figure S4**

*Correlation between Emotion Classification Performance and self-rated Music Skills (GOLD-MSI)*

**

*Note. The x-axis shows the different subscores of the Gold-MSI (Active, Perception, Singing, Emotion, and Education) as well as the General Music Education score (General ME). The y-axis shows the vocal emotion classification performance separately for each Morph Type (Full, F0 and Timbre) and averaged across Morph Types (Averaged). p-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995; false discovery rate set to 0.05, total number of tests = 24)*

**Table S7 – PROMS and VER, correlations**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PROMSAvg | Pitch | Melody | Timbre | Rhythm |
| VERAvg | **0.44 (< 0.001)** | 0.22 (0.09) | **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. This table is identical with Table 2 from the manuscript. VER = Vocal Emotion Recognition performance. p-values of Tables S7 – S15 were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995)*

**Table S8 – PROMS and VER, controlled for musical training**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PROMSAvg | Pitch | Melody | Timbre | Rhythm |
| VERAvg | **0.41 (0.005)** | 0.13 (0.369) | **0.35 (0.009)** | 0.18 (0.203) | **0.32 (0.014)** |
| Full-Morphs | **0.39 (0.005)** | 0.15 (0.322) | **0.36 (0.008)** | 0.24 (0.081) | 0.22 (0.111) |
| F0-Morphs | **0.36 (0.008)** | 0.04 (0.739) | **0.31 (0.014)** | 0.13 (0.369) | **0.33 (0.011)** |
| Timbre-Morphs | 0.13 (0.369) | 0.09 (0.502) | 0.06 (0.643) | 0.06 (0.643) | 0.12 (0.369) |

*Note. VER = Vocal Emotion Recognition performance.*

**Table S9 – PROMS and VER, logistic regressions**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PROMSAvg | Pitch | Melody | Timbre | Rhythm |
| VERAvg | **1.37 (< 0.001)** | 0.6 (0.064) | **0.73 (0.001)** | 0.54 (0.053) | **0.87 (0.005)** |
| Full-Morphs | **2.61 (< 0.001)** | **1.27 (0.03)** | **1.46 (< 0.001)** | **1.13 (0.026)** | **1.21 (0.03)** |
| F0-Morphs | **1.52 (0.005)** | 0.43 (0.314) | **0.83 (0.006)** | 0.55 (0.157) | **1.2 (0.005)** |
| Timbre-Morphs | 0.58 (0.209) | 0.39 (0.314) | 0.21 (0.407) | 0.2 (0.51) | 0.48 (0.209) |

*Note. Estimated parameters of the logistic regression with the syntax VER ~ PROMS.*

**Table S10 – PROMS and VER, logistic regressions, controlled for musical training**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PROMSAvg | Pitch | Melody | Timbre | Rhythm |
| VERAvg | **1.62 (0.008)** | 0.39 (0.322) | **0.85 (0.01)** | 0.43 (0.163) | **0.77 (0.019)** |
| Full-Morphs | **2.66 (0.01)** | 0.67 (0.322) | **1.50 (0.01)** | 0.85 (0.134) | 0.90 (0.142) |
| F0-Morphs | **1.83 (0.014)** | 0.09 (0.852) | **1.02 (0.019)** | 0.42 (0.312) | **1.12 (0.014)** |
| Timbre-Morphs | 1.00 (0.142) | 0.56 (0.287) | 0.38 (0.322) | 0.21 (0.541) | 0.51 (0.232) |

*Note. Estimated parameters of the logistic regression with the syntax VER ~ PROMS + musical training.*

**Table S11 – MSI and VER, correlations**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | General Sophistication | Active  Engagement | Musical Training | Emotions | Singing Abilities | Perceptual Abilities |
| VERAvg | **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 (0.12) | 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. This table is identical with Table 3 from the manuscript. VER = Vocal Emotion Recognition performance.*

**Table S12 – MSI and VER, controlled for musical training**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | General Sophistication | Active  Engagement | Emotions | Singing Abilities | Perceptual Abilities |
| VERAvg | 0.28 (0.093) | 0.08 (0.718) | -0.1 (0.649) | 0.23 (0.206) | 0.20 (0.22) |
| Full-Morphs | **0.35 (0.039)** | 0.11 (0.604) | -0.05 (0.791) | 0.31 (0.061) | 0.20 (0.22) |
| F0-Morphs | 0.22 (0.206) | 0.02 (0.863) | -0.12 (0.593) | 0.21 (0.216) | 0.16 (0.368) |
| Timbre-Morphs | 0.07 (0.737) | 0.04 (0.84) | -0.05 (0.791) | 0.03 (0.863) | 0.08 (0.718) |

*Note. VER = Vocal Emotion Recognition performance.*

**Table S13 – MSI and VER, logistic regressions**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | General Sophistication | Active  Engagement | Musical Training | Emotions | Singing Abilities | Perceptual Abilities |
| VERAvg | **0.04 (0.024)** | 0.03 (0.176) | 0.02 (0.096) | 0.01 (0.652) | **0.04 (0.01)** | **0.05 (0.009)** |
| Full-Morphs | **0.08 (0.006)** | 0.06 (0.059) | **0.05 (0.024)** | 0.04 (0.374) | **0.09 (0.005)** | **0.09 (0.006)** |
| F0-Morphs | 0.04 (0.096) | 0.02 (0.524) | 0.02 (0.191) | -0.01 (0.793) | 0.04 (0.059) | 0.05 (0.064) |
| Timbre-Morphs | 0.01 (0.558) | 0.01 (0.65) | 0 (0.917) | 0.01 (0.741) | 0.02 (0.515) | 0.02 (0.347) |

*Note. Estimated parameters of the logistic regression with the syntax VER ~ MSI.*

**Table S14 – MSI and VER, logistic regressions, controlled for musical training**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | General Sophistication | Active  Engagement | Emotions | Singing Abilities | Perceptual Abilities |
| VERAvg | 0.08 (0.11) | 0.01 (0.841) | -0.01 (0.783) | 0.05 (0.102) | 0.05 (0.102) |
| Full-Morphs | 0.13 (0.099) | 0.01 (0.816) | -0.01 (0.841) | 0.1 (0.102) | 0.08 (0.099) |
| F0-Morphs | 0.06 (0.313) | -0.02 (0.783) | -0.04 (0.313) | 0.05 (0.219) | 0.05 (0.217) |
| Timbre-Morphs | 0.07 (0.217) | 0.02 (0.598) | 0.01 (0.798) | 0.04 (0.313) | 0.05 (0.217) |

*Note. Estimated parameters of the logistic regression with the syntax VER ~ MSI + musical training.*

**Table S15 – MSI and PROMS, correlations**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | General Sophistication | Active  Engagement | Musical Training | Emotions | Singing Abilities | Perceptual Abilities |
| PROMSAvg | **0.69 (<.001)** | **0.53 (<.001)** | **0.61 (<.001)** | 0.21 (0.067) | **0.66 (<.001)** | **0.69 (<.001)** |
| Pitch | **0.59 (<.001)** | **0.45 (<.001)** | **0.56 (<.001)** | **0.41 (<.001)** | **0.56 (<.001)** | **0.61 (<.001)** |
| Melody | **0.70 (<.001)** | **0.58 (<.001)** | **0.62 (<.001)** | **0.31 (0.009)** | **0.66 (<.001)** | **0.68 (<.001)** |
| Timbre | **0.32 (0.007)** | **0.25 (0.033)** | **0.24 (0.041)** | -0.06 (0.595) | **0.37 (0.002)** | **0.33 (0.005)** |
| Rhythm | **0.27 (0.023)** | 0.18 (0.129) | **0.25 (0.035)** | -0.09 (0.475) | **0.26 (0.027)** | **0.31 (0.008)** |