1 Cognitive Music Listening Space: A Multivariate Approach

2 Brendon Mizener1, Mathilde Vandenberghe2, Hervé Abdi1, & Sylvie Chollet2

3 1 University of Texas at Dallas

4 2 YNCREA

5 Author Note

6 Add complete departmental aﬀiliations for each author here. Each new line herein

7 must be indented, like this line.

8 Enter author note here.

9 The authors made the following contributions. Brendon Mizener: Stimuli creation,

10 Survey design & creation, Data collection & processing, Statistical analyses, Writing -

11 Original draft preparation; Mathilde Vandenberghe: Original concept, Survey design &

12 creation; Hervé Abdi: Writing - Review & Editing, Statistical guidance; Sylvie Chollet:

13 Original concept.

14 Correspondence concerning this article should be addressed to Brendon Mizener, 800

15 W. Campbell Rd., Richardson Tex. E-mail: [bmizener@utdallas.edu](mailto:bmizener@utdallas.edu)

16 Abstract

17 French and American participants listened to new music stimuli and

18 evaluated these stimuli using either adjectives or quantitative musical dimensions.

19 Results were analyzed using correspondence analysis (CA), Hierarchical cluster analysis

20 (HCA), Multiple Factor Analysis (MFA), and Partial Least Squares Correlation (PLSC).

21 All except the HCA used Bootstrapping and Permutation testing for inferences.

French and American listeners differed when they described the musical stimuli

23 using adjectives, but not when using the quantitative dimensions. We did not control how

24 participants listened to the stimuli, but they were encouraged to use headphones or listen

25 in a quiet listening environment. Participants were also able to complete the survey using a

26 mobile device. The present work serves as a case study in research methodology that allows for a

27 balance between relaxing experimental control and maintaining statistical rigor.

28 *Keywords:* Music, Emotion, Multivariate Analyses

29 Word count: 5631

30 Cognitive Music Listening Space: A Multivariate Approach

31

32 World events over the last year have demonstrated the need for an expansion of

33 traditional experimental paradigms and robust

34 and consistent remote or online data collection. However, that shift in collection

35 necessitates a consequent shift in analysis. Experiments conducted in labs are

well controlled and so provide clean data compared to data

37 collected using online surveys. Dirtier data means that most likely, some

38 of the assumptions associated with traditional univariate analyses, hypothesis testing, and

39 inferences are violated and therefore these data require specific methods of analysis and inference.

40 Here we present a case study using real data that were obtained using online multinational data

41 collection and were analyzed using multivariate analyses.

The initial motivation for the present work came from a study

42 investigating cross modal sensory mapping between flavor perception (specifically of beer)

43 and music perception. As such, this study was designed to investigate whether a music

44 cognitive listening space could be established using the experimental and analysis paradigm

45 outlined below, to allow cross-modal comparison. Additional questions arise from the study

46 itself: Are there significant differences in how participants from different nationalities (and

47 by extension musical cultures) perceive and describe music? Are there

48 parallels in how music is evaluated using music non-specific descriptors and music-specific

49 qualities?

50 Noise in online data collection comes in many forms, including

51 incomplete responses, environment, or technology used to access the survey. Maintaining

52 experimental rigor through these sources of variance can be diﬀicult, but is not

53 unmanageable. The check-all-that-apply (CATA) method (Meyners & Castura, 2014) is an example of

54 a data collection technique that features a number of benefits in this regard. Other sources

55 of noise can be minimized by increasing sample size—a procedure relatively easy when collecting

56 online data—and by using multidimensional statistical methods.

The CATA method is used to measure how participants evaluate a set of stimuli presented one at a time. For each stimulus, participants are shown a list of descriptors and are asked

59 to select the descriptors that apply to the presented stimulus. This mode of presentation minimizes participant cognitive demand by providing a rapid means of assessing sensory

61 profiles (Ares et al., 2010; Meyners & Castura, 2014). Katz and Braly (1933) provides an

62 early example of the use of the CATA paradigm in the psychological sciences. It is not

63 terribly common in the psychological sciences anymore, but has been and continues to be

64 used widely in sensory evaluation (Abdi & Williams, 2010). A single stimulus may be

65 described by multiple adjectives, so selecting only one ‘correct’ answer is not necessary.

66 Similarly, the adjectives that may only partially describe the stimulus, or do so

67 tangentially, are likely to be selected by fewer participants, and adjectives that more

68 completely describe the stimulus will be selected by more participants. Thus we have a

69 data collection paradigm that allows for a gradient across the adjectives and stimuli that is

70 robust to violations, either intentional or not. A more complete treatment of the value of

71 such a data collection mechanism, including assessments in which there is a ‘correct’

72 answer, is found in Coombs et al. (1956). \*\*\* the last part of the paragraph “rambles” a bit and I am sure that I get that we want to say here. At that part of the introduction, readers are looking for a statement of the problem that we are going to tackle, but currently they will have a hard time finding it. Maybe we need to find some time to go through the intro to thibk about it. \*\*\*

73 Multivariate analyses are useful tools for dealing with ‘dirty’ data, that is, data with

74 a smaller signal-to-noise ratio. Univariate analyses are less than ideal for studies run online

75 because any violations in the one target variable reduce the signal, and make it more

76 diﬀicult to interpret results and draw conclusions. One solution is greater power, another is

77 to increase the number of variables and change the analytical paradigm. Using a

78 multivariate perspective helps the analysis. In a solution to a system in which there are ten

79 or more dimensions, greater noise in one or two of those dimensions is less intrusive

80 because the multivariate solution evaluates the total variance in all of the dimensions,

81 instead of the variance for each individual dimension separately. This makes the system

82 and the solution more robust to violations and noise. Additionally, the robustness of this

83 type of analysis is compounded by greater power.

### 84 Music Perception

85 Quantifying music perception is an interesting problem that gets at the heart of this

86 specific issue. Music is an artistic and communicative acoustic medium that unfolds over

87 time. Most music studies impose strict controls over participants’ listening environment to

88 minimize differences in the auditory signal and environment. Small changes can affect

89 listeners’ perception, especially when the study involves timing or specific tuning. However,

90 the experimental controls may be loosened slightly when investigating holistic music

91 listening, as the macro signal is more important than any individual facet.

92 In this holistic listening paradigm, listeners continuously evaluate incoming

93 information and compare it with that which came before. These comparisons are related to

94 both technical and affective aspects of music. While these two aspects of music are

95 theoretically distinct, in practice there is a great deal of interplay between the two.

96 Listeners respond affectively to technical aspects of music, and composers use various

97 musical and compositional techniques things to reflect the internal emotional states they

98 want to express. And, although isolated musical characteristics have been demonstrated to

99 have a certain effect on listeners’ affective perception (Bruner II, 1990), the interactions

100 between multiple musical characteristics provide a more complicated challenge, to say

101 nothing of the individual associations that participants bring to the table (Kopacz, 2005).

102 One of the reasons these interactions have been diﬀicult to pin down is that models

103 like ANOVA which use only a few variables are limited by how many variables a researcher

104 can include while remaining coherent. Thus, the many studies that use strict controls and

105 vary only one element of music at a time to evaluate how various technical aspects of music

106 correspond to emotions for the purpose of induction, (see Bruner II (1990) for a summary)

107 do not reflect the complexity inherent to music and music listening.

108 Research on music and emotion is a similarly well-trod topic. See, for example, Juslin

109 and Sloboda (2010). An early study by Wedin (1969) supported Osgood’s (1955) theory

110 that valence and arousal were the two most salient dimensions in evaluating emotionally

111 charged stimuli, including music. Studies supporting the existence of the valence-arousal

112 plane (Osgood & Suci, 1955) have replicated these results many times. In fact, recent

113 trends in experimental procedure in behavioral studies of music and emotion have been for

114 participants to rate music using arousal and valence sliders (Bigand et al., 2005; Madsen,

115 1997), specifically asking the participants to rate on those two dimensions. This is useful,

116 but limiting, as it provides fine-grained detail on the level of arousal or valence a given

117 stimulus provides, but does not qualify that information. There have been a few studies

118 that have specifically investigated dimensions beyond those first two (for example Rodà et

119 al. (2014)), and recent theories of the dimensionality of emotion include as many as 27

120 dimensions (Cowen & Keltner, 2017), but the various results on perceptual dimensions

121 beyond valence and arousal are inconclusive.

122 One common analysis used for these kinds of studies is Multidimensional Scaling

123 (MDS). MDS was introduced fairly early on as a means of evaluating the perceptual space

124 around musical excerpts (Wedin, 1969, 1972). Studies in this vein have continued to date.

125 However, MDS is primarily a distance analysis, and is therefore limited in the perspective

126 it can provide. It is commonly used to represent the cognitive distance between stimuli.

127 This is an interesting application of this analysis, but doesn’t use it to its full potential. We

128 suggest that this analysis may be more effective in representing the cognitive differences in

129 the behavior of participants.

### 130 Present questions & methods of analysis

131 In this study, we attempt to address three specific issues with the field as a whole:

132 mode of investigation, sample & size, and analysis. The basic question was simple: how do

133 French and American participants describe music? Our investigative paradigm, along with

134 sample and size, are addressed in the methods section below, but we felt it may be useful

135 to provide a quick overview of the analytical techniques for readers who may be unfamiliar.

136 **Correspondence Analysis.** The primary analysis used on the data collected in

137 the surveys is Correspondence Analysis (CA). CA has many names, and has been

138 ‘discovered’ many times by many people. There are a number of excellent references that

139 illustrate the calculative (Greenacre, 1984) and graphical or geometrical (Benzécri, 1973).

140 CA is similar to Principal Components Analysis (PCA), except that it allows for the

141 analysis of qualitative data. Data for a CA is organized in a contingency table or a pseudo

142 contingency table. A contingency table is be when a participant selects only one option

143 from a list for each stimulus, resulting in a table for each participant with one and only one

144 one (1) per row, and a pseudo contingency table has as many ones as items selected for a

145 given stimulus. Because we use a CATA paradigm for the adjective survey, we use the

146 latter. In this table, the value in a given cell represents the relationship between the

147 observation and the variable symmetrically, that is, it is both the number of times a

148 variable was selected to be associated with an observation, and the number of times an

149 observation was selected to be associated with a variable. Because of this, the variance of

150 the table as a whole can represent either the variance associated with the rows or the

151 columns, depending on how it is analyzed. Thus, this technique allows us to plot factor

152 scores for both rows and columns in a single space. In addition to the standard factor plots,

153 we used permutation tests and bootstrapping to make inferences.

154 **Partial Least Squares Correlation.** Partial Least Squares Correlation (PLSC)

155 (Abdi & Williams, 2013) analyzes two data tables that have the same information either on

156 the observations (rows) or variables (columns). The PLSC extracts the covariance between

157 two tables in the form of *latent variables*. This technique is commonly used in

158 neuroimaging studies to evaluate correlations between matrices of imaging data and of

159 behavioral or task data (Krishnan et al., 2011). In our context, the PLSC extracts the

160 information that is shared between the adjectives ratings and the musical dimensions

161 ratings. The stimuli are on the observations (rows) for both data tables. Additionally, the

162 contributions and loadings will show us which variables are responsible for creating or

163 defining the primary axes of similarity between the two data sets. There are some criticisms

164 of this technique that argue that it is overpowered, that it can ‘find’ spurious correlations,

165 and to that end we would simply suggest caution when interpreting PLSC results.

166 **Multidimensional Scaling.** Multidimensional Scaling (MDS) (Borg & Groenen,

167 2005) analyzes a square, symmetrical distance matrix in which each cell represents the

168 distance, or the amount of difference, between the item on the row and on the column. The

169 resultant factor scores are the relative distance between all of the points, and are plotted

170 similarly to PCA. In this case, we calculated a symmetrical distance matrix for the

171 participants, to see whether there were any significant differences between groups of

172 participants when grouped according to any of the factors extracted from the demographics

173 survey.

174 **Multiple Factor Analysis.** Multiple Factor Analysis (MFA) is the only

175 unplanned analysis used in this study, and is also the newest (Abdi et al., 2013). We chose

176 to run this analysis post hoc after finding significant mean differences between French and

177 American participants for one of the surveys. MFA is uniquely suited to analyze and

178 visualize the relative contributions of multiple tables or groups of variables simultaneously,

179 and allows for the disambiguation of the various contributions of either a population or a

180 set of variables in a plot. The observations must all be the same for MFA, but analysis can

181 either evaluate the entire population, with the variables grouped in ways that are useful or

182 valuable to isolate, or with separate populations, using all the same variables for both

183 groups. The number of tables (i.e., populations or groups of variables) you choose to

184 analyse is limited by what makes sense, either mathematically by way of planned analyses

185 or visually in the partial factor scores plots. In any case, the visualization output for this

186 plot provides the researcher with factor scores of the observations overall, and partial factor

187 scores showing how each of the tables contributed to each observation; where each

188 individual weighted table would fall in the factor space relative to the other/s. Because the

189 tables for this analysis are weighted according to their overall inertia, with larger tables

190 being weighted less than smaller tables, this is a very useful technique when dealing with

191 unbalanced groups.

192 **Inference Methods.** Because the methods outlined above are not inferential

193 methods, and do not inherently allow for hypothesis testing, we need to also apply methods

194 that help with that. To acheive this, we use permutation testing (Berry et al., 2011) and

195 bootstrapping (Hesterberg, 2011).

196 Permutation testing shuffles the data and recomputes the eigenvalues for each

197 iteration. Because the eigenvalues extracted from these data tables are also an indication of

198 how much variance is extracted by each dimension, random data should give us smaller

199 eigenvalues, indicating a weaker signal. Therefore, if the observed eigenvalues are larger

200 than a certain threshold, we can infer that the data we collected do, in fact, represent

201 something real or important. Importantly, this is determined by the number of iterations

202 that we permute, we can only infer to that degree. If we want to infer to the standard

203 alpha level of .05, then we would need to run at least 100 permutations, and hope that the

204 observed result was one of the largest five values.

205 Bootstrapping, on the other hand, is resampling with replacement. We use this

206 technique for two reasons: the first is to resample the factor scores to establish a confidence

207 interval around the mean of the groups, the other is to resample with a focus on the

208 loadings, to see which of the observations and variables load consistently on the dimensions

209 we’re interpreting. Both give us an idea of the consistency of the data, and can once again

210 give us an idea of the statistical significance of mean differences based on the number of

211 iterations performed.

212 **Methods**

213 **Participants**

214 Participants (N = 604) were recruited similarly for both Experiments 1 and 2, and

215 thus are discussed simultaneously here. Participants for this study were recruited in

216 multiple ways. The participants in the United States (n = 292) were recruited using the

217 traditional method of offering experimental participation credit, and also via social media.

218 French participants (n = 312) were recruited by word of mouth, email, and social media.

219 The only restrictions on participation were that the participant must have self-reported

220 normal hearing. We recognize that although we suggest that data collected in this way

221 have a much greater hypothetical reach, the data here represent a) a convenience sample,

222 b) that is limited to participants that have access to the internet, and c) because of the

223 nature of social media, many of the participants in the researchers’ social circles are

224 themselves students, thus providing an additional confound. However, these specific

225 limitations could be remedied when designing and implementing future research.

226 The population we recruited was different for the two experiments. For Experiment 1,

227 we specifically sought out highly trained musicians (n = 84) with ten years or more of

228 music training. We recruited this population for two reasons: firstly, as a validation step,

229 to ascertain whether the stimuli truly reflected the composer’s intent. Secondly, we had the

230 goal of evaluating the perceptual effect of the stimuli as it relates specifically to the musical

231 qualities. These perceptual evaluations were to then be correlated with the adjectives

232 selected by those who participated in the adjectives survey. Participants were recruited for

233 Experiment 2 (n = 520) without regard to level of music training.

234 Of the responses to Experiment 1, 51 were removed to incomplete data (nF = 45, nA

235 = 6), leaving a total of 33 for the analysis. Of the responses to Experiment 2, 160 were

236 removed for not completing the survey (nF = 140, nA = 20), leaving a total of 360. Of the

237 responses to the survey administered in the US, participants were excluded from analysis if

238 they indicated a nationality other than American. “Asian-American,” for example, was

239 included, but “Ghanian” was not. This left a total of 279 survey responses for Experiment

240 2 and 312 for analysis across both experiments.

241 All recruitment measures were approved by the UT Dallas IRB.

242 **Material**

243 **Stimuli.** All stimuli were original, novel musical excerpts, in various western styles,

244 composed for this study. They were designed to evaluate a number of musical dimensions

245 and control for others (e.g., timbre). The stimuli were all string quartets, in order to

246 control for the confounding factor that different instruments are fundamentally described

247 in different ways. All stimuli were between 27s and 40s long, with an average length of

248 32.4s. The intent was to have all stimuli be around 30s long while preserving musical

249 integrity. All stimuli were composed using finale version 25.5.0.290 [cite finale] between

250 April 13 and June 18, 2020. Stimuli were recorded as wav files directly from finale using

251 the human playback engine and embedded into each question in qualtrics in that format.

252 **Surveys.** There were two separate surveys presented to participants. The survey

253 used in Experiment 1 (hereafter: Qualities Survey/QS) evaluated the musical stimuli on

254 concrete musical qualities like meter and genre. The survey used in Experiment 2

255 (hereafter: Adjectives Survey/AS) asked participants to evaluate the stimuli using

256 adjectives using the CATA paradigm. Both surveys also captured participants’

257 demographic data, including age, gender, nationality, occupation, and musical experience.

258 The qualities assessed in the QS were selected from standard music-theoretical

259 descriptors of western music. For example, when rating the excerpts on tempo, participants

260 were asked to rate the excerpt using the scale *Very Slow*, *Slow*, *Moderately Slow*, *Moderate*,

261 *Moderately Fast*, *Fast*, and *Very Fast*. The full list of musical qualities and answer choices

262 is listed in the supplementary materials. The words for the AS were selected using

263 Wallmark (2019) as a guide and in consult with a French professional musician. Some

264 words were initially selected in French and some in English. In all cases, words were

265 selected for which there was a clear French (vis-à-vis English) translation. The words are

266 listed in English and in French in the supplementary materials.

267 **Procedure**

268 Participants were provided with a link to either the AS or the QS. Both surveys were

269 administered using Qualtrics. After standard informed consent, participants listened to 15

270 excerpts and answered questions. Participants were instructed to listen to the excerpts

271 presented either using headphones or in a quiet listening environment, but that was not

272 strictly controlled, nor was it part of the survey. Participants in Experiment 1 answered 10

273 questions per excerpt, rating the excerpts using the qualities and scales provided.

274 Participants in Experiment 2 answered a single question per excerpt, in which they selected

275 any and all adjectives that they felt described the excerpt. Demographic survey questions

276 followed the experimental task.

277 **Data Processing.** Raw data were cleaned and processed in Excel and R. This

278 included translating all French responses to English for ease of analysis. Data were cleaned

279 and transformed into a pseudo contingency table for each participant, with the stimuli, as

280 observations, on the rows and the responses as variables on the columns. In these

281 individual tables, a one (1) at the intersection of each row or column indicates that the

282 participant selected that adjective or musical quality for that stimulus. A zero means that

283 they did not. These individual tables were all compiled into into two ‘bricks,’ or

284 three-dimensional arrays of data with the same structure for the rows and columns, and

285 the participants on the third dimension, which we will refer to as ‘pages’ here. Each array

286 was then summed across pages into a single, two dimensional, summary

287 pseudo-contingency table, so that any given cell contained the total number of times a

288 participant selected a given adjective or quality for a given stimulus.

289 Since we did not use *a priori* grouping variables for the excerpts or adjectives, the

290 summed tables were evaluated using hierarchical cluster analyses to see what groupings

291 arose during evaluation. Hierarchical cluster analyses, included in supplementary materials,

292 captured groupings of the excerpts when rated by the adjectives and when rated on musical

293 qualities. The musical qualities were grouped by quality (e.g., levels of tempo, types of

294 genre). These groupings were used for coloring on the plots and for statistical inferences.

295 **Results**

### 296 Experiment 1: Musical Qualities Survey

297

298

299

300

301

302

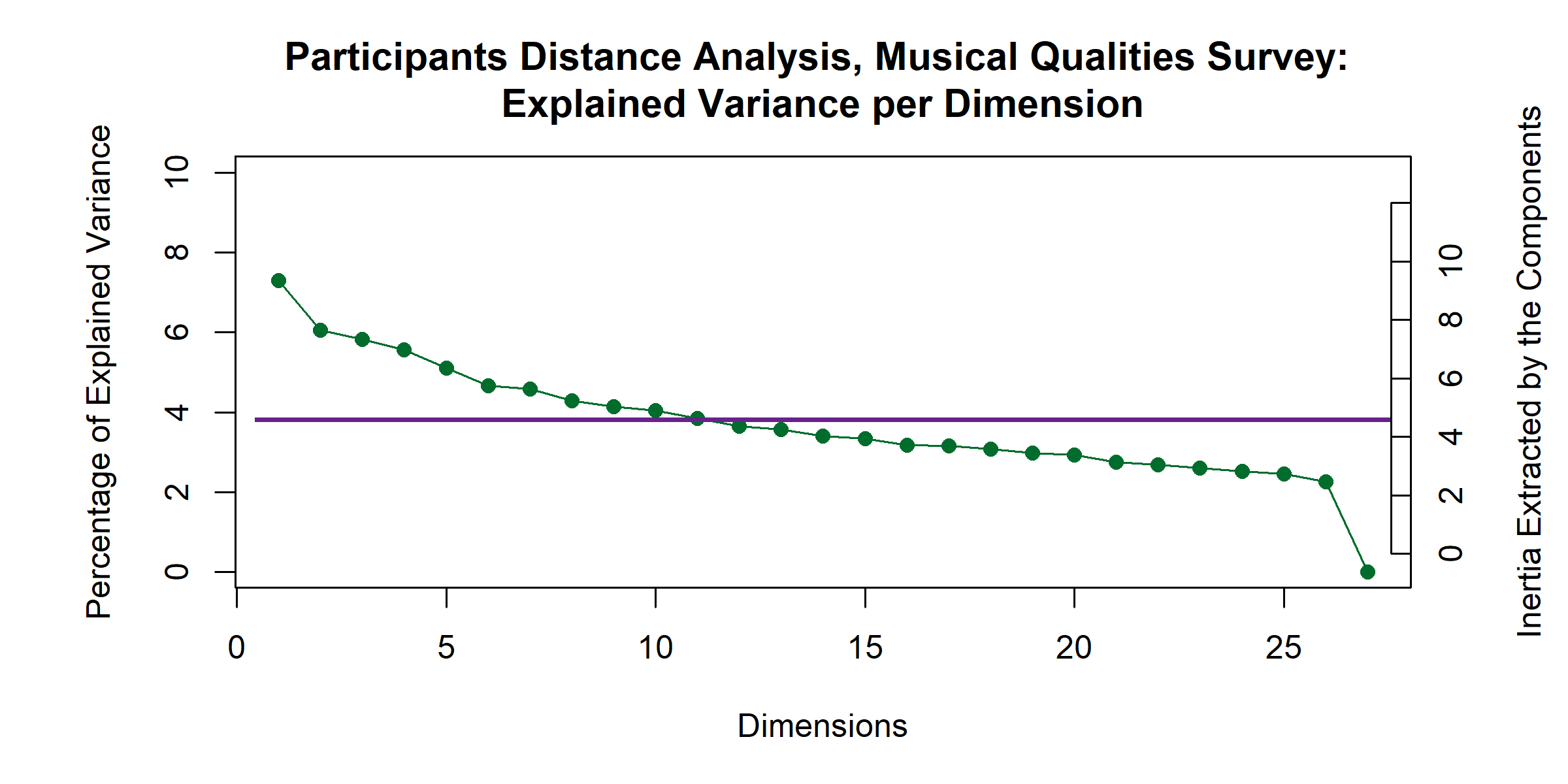
303

**Participants.** The

scree plot in Figure [1](#_bookmark0) shows the eigenvalues for the distance analysis between musical experts. The usual guideline of analyzing only dimensions with eigenvalues greater than one seems prohibitive here, as all dimensions except the last have > 1. For

𝜆

to focus on the first two dimensions, with 𝜆 𝜆



*Figure 1*

304 the purposes of this case study, we’ve opted

305 = 9.06 and = 7.52, respectively. This scree

306 plot suggests that each of the participants is contributing similarly to the dimensionality of

307 this analysis. To evaluate this, we ran a Multidimensional Scaling (MDS) analysis on a

308 double-centered cross product symmetric distance matrix calculated from the pages of the

309 brick. This analysis revealed no significant difference between the experts based on any of

310 the grouping variables used. The factor plots in Figure [2](#_bookmark1) show how the means of the factor

311 scores, grouped by nationality and gender identity, respectively, show the means clustered

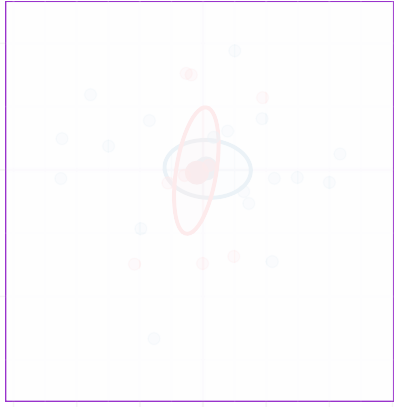
312 on top of one another, right at the origin. The overlapping ellipses are the confidence

313 intervals for the means.

*Factor Scores for Expert Ratings*

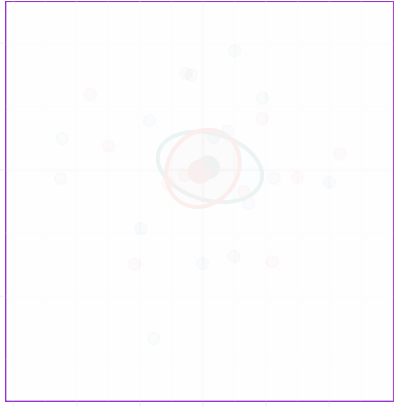
Colored according to Nationality Colored according to Gender

1 1



**AM**

**FR**



**F**

**M**

Dimension 2. λ = 7.52. τ = 6%

Dimension 2. λ = 7.52. τ = 6%

0 0

−1 −1

−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5

Dimension 1. λ = 9.057. τ = 7%

−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5

Dimension 1. λ = 9.057. τ = 7%

*Figure 2*

314

315

316

317

318

319

320

321

322

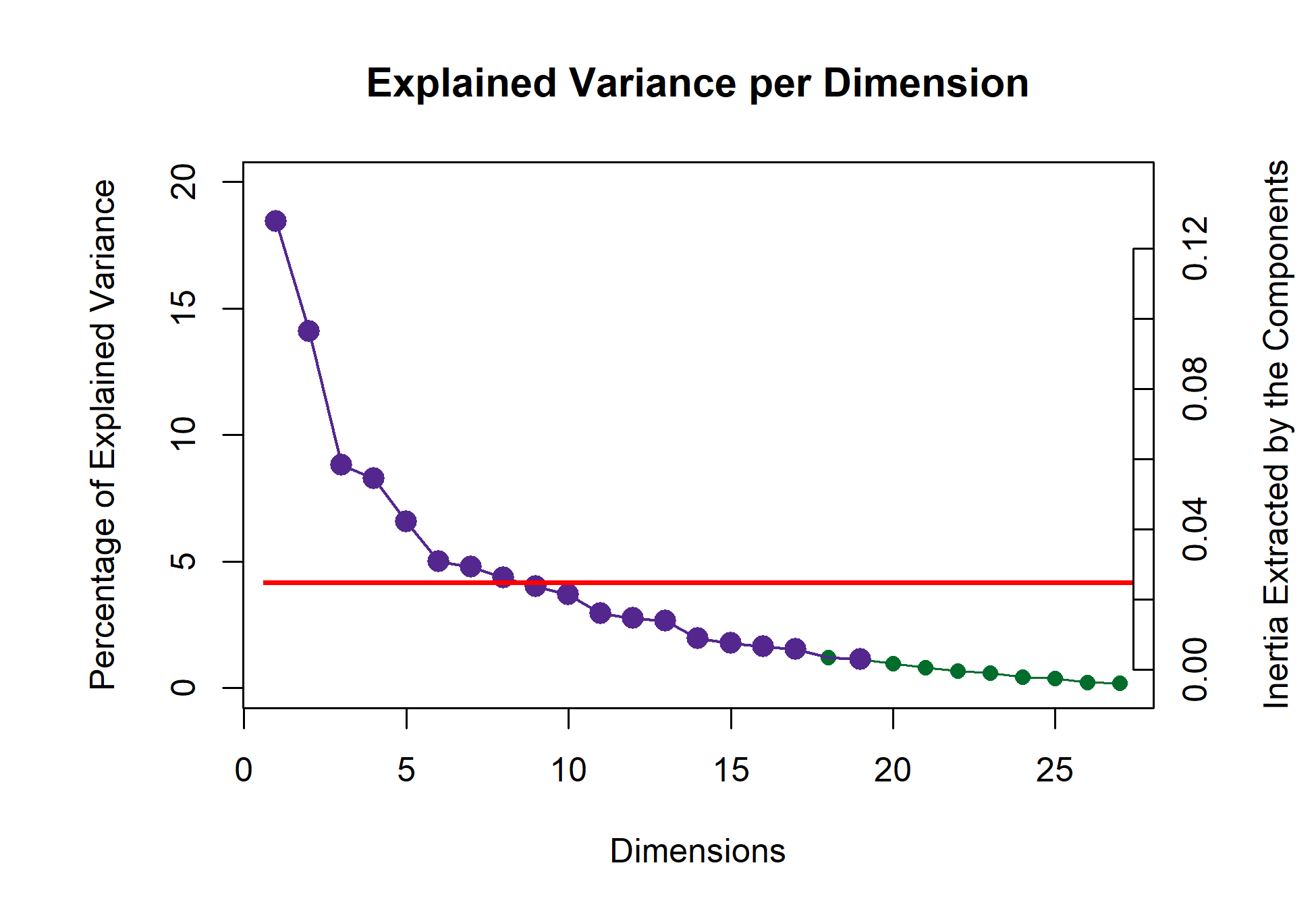
323

**Excerpts.** The

scree plot for the analysis of the musical quality ratings survey, Figure [3](#_bookmark2), shows

the high dimensionality of this space, with the first three dimensions extracting a total of 18.44%, 14.09% and 8.81% respectively, totaling only 41.34% of the variance.

It isn’t until we get to the 11th dimension that we see >80% of the variance explained. However, given that the assumption in an



*Figure 3*

324 analysis like this is that the sample is random, it’s important to take these numbers with a

325 grain of salt. Music itself is not random, and in a single excerpt of music of the type that

326 was presented in this study, repetition is common, and some musical qualities are

327 inextricably linked, for example some stylistic elements with genre. Graphing the variable

328 loadings (see Figure [4](#_bookmark3)) of the musical qualities shows which ones contribute the most to the

329 first two dimensions. Because of how CA is calculated, we know that the excerpts that load

330 on the same dimension and direction as the musical qualities are the excerpts that are most

331 associated with those qualities. The contributions shown here are only those that

332 contribute significantly to the first two dimensions. There are some obvious groups of

333 variables, especially tempo and articulation in the first dimension, with fewer contributions

334 from the dynamics group. The tempo variables, which are a continuum, load from high

335 (tempo.F6 and tempo.F7) in the positive direction to low (tempo.F2 and tempo.F1) in the

336 negative direction. Other contributions are one-off: major harmony, triple meter, classical

337 genre, undulating contour, and disjunct motion. The excerpts that load positively, and are

338 therefore associated with the qualities that load in the positive direction, are all from group

339 2: Excerpts 4, 13, 23, and 26. The ones that load in the negative direction are from mostly

340 from group 4: Excerpts 7, 10, 24, and 27, with one from group 3, Excerpt 3.

341 The second dimension seems to dominated by a few groups: harmony, meter, genre,

342 dynamics. The one-offs are slow tempo, ascending contour, and “no melody.” The excerpts

343 that load significantly on this dimension are from all four groups. In the positive direction,

344 it’s Excerpts 7, 12, 15, and 27 from Group 4, and Excerpt 19 from Group 1. In the

345 negative direction it’s Excerpts 2, 3, 11, and 17. All are from group 3 except for Excerpt 2,

346 which is from Group 2. A full enumeration of contributions, loadings, and boostrap ratios

347 is available at the github url in the author note.

348 **Discussion.** The graph depicted in Figure [5](#_bookmark5) is a biplot depicting how excerpts and

349 variables plot in the same space. This biplot is possible because of the nature of

350 correspondence analysis. Because the rows and columns of the contingency table X by

351 definition have the same variance, the eigenvalues extracted from any matrix *X* are the

⊤

352 same as *X* . Thus the axes on which the factor scores are plotted are the same for both the

353 rows and the columns. However, interpretation requires some discernment. The distance

354 between the excerpts can be interpreted directly as similarity, and the distance between the

0.15

0.10

Contributions

Excerpt.3

0.05

0.00

*Contributions*

0.15



motion.Disjunct

0.10

Excerpt.24

Excerpt.27

Contributions

tempo.F3

tempo.F2

tempo.F1

meter.Trip

dyn.Soft

artic.Legato

0.05

0.00

Excerpt.26

harm.Major

tempo.F5

tempo.F6

tempo.F7

nre.Classical

dyn.Loud

t.Undulating

artic.Marcato

rtic.Staccato

−0.05

Excerpt.7

Excerpt.10

−0.05

−0.10

Excerpt.4

−0.10

0.1

Excerpt.11

Excerpt.13

Excerpt.23

Excerpt.17

harm.Minor

meter.Trip

genre.Clas

genre.Rom

genre.Baro

dyn.Soft

0.1

Excerpt.2

Excerpt.3

0.0

Contributions

Contributions

harm.Quin harm.Blues

tempo.F1

meter.Quad

enre.Contemporary

genre.Modern

genre.Jazz/Blues

dyn.Loud dyn.Gra\_Decr

cont.Ascending

melody.No

0.0

Excerpt.7

Excerpt.12

Excerpt.15

Excerpt.19

Excerpt.27

−0.1

−0.1

−0.2

*Figure 4*

355 musical qualities can be interpreted directly as similarity, but the distance between a

356 quality and an excerpt cannot. Instead, the angle between an excerpt and a quality is

357 indicative of their correlation. An angle of 0 indicates a correlation of 1, an angle of 90

358 indicates a correlation of 0, and an angle of 180 indicates a correlation of -1.

359 Overall, this helps us to evaluate what contribute to the excerpt groupings. These

360 first two dimensions suggest that the hierarchical cluster analysis (see supplementary

361 materials) revealed groupings roughly according to genre. However, there are two notable

362 outliers. Excerpts 6 and 14 are unique in that they are each the only representative of their

363 respective genres. Excerpt 6 is minimalist, à la Steve Reich, and Excerpt 14 is jazzy.

364 Preliminary versions of this analysis showed that they dominated the 2nd and 3rd

365 dimensions, respectively (see supplementary materials for visualizations). In the plot below,

366 they are included instead as supplementary projections, essentially ‘out of sample’ elements.

367 Their placement on the plot below alludes to the fact that the dimensionality of this space

368 may in fact be related to musical genre or family. Although they dominated the space

369 when included in the sample, they are much closer to the barycenter of the plot when

370 included as out of sample. Were they to fall exactly on the origin, that would suggest that

371 they shared no information whatsoever with the other excerpts included in the analysis.

372 The disparity between their placement on the graph below and their placement on the

373 graphs in which they are included in the main sample suggests that they share some

374 information, but there is still a large amount of information that is not accounted for in the

375 factor space depicted in Figure [5](#_bookmark5).

376 One perceptual element that is revealed here is that tempo and dynamics seem to

377 contribute, intensity-wise, similarly to the first dimension. This points to two specific

378 things. Firstly, it highlights possible bias in the compositional process. The excerpts were

379 not intentionally composed with those characteristics being similar in mind, but it’s

380 entirely possible that the high or low arousal levels of the various excerpts that participants

381 respond to also drove some of the compositional process, and that turned up in the results.

382 Secondly, it’s possible that the level of arousal was conflated between various musical

383 qualities. For example, the intensity and therefore tempo of a stimulus may have been

384 affected by the volume or dynamics (**Kamenetsky1997?**). Perception of tempo is also

385 affected by note rate or event density, which is also tied to arousal. In two pieces played at

386 the same tempo, the one with more notes per unit time is more likely to be judged faster

387 than one with fewer (**Drake1999?**). There are also a few musical elements revealed from

388 the associations. The term staccato means short, or light and separated, and the term

389 legato means smooth and connected. The participants in this experiment didn’t have access

390 to the notation, so they would be judging the excerpts aurally only. Between faster and

391 slower excerpts, notes of the same rhythmic value take up less time in the faster excerpts,

392 and may be more likely to be judged as light and separate, regardless of what the actual

393 articulation was. Slow tempo and legato are associated differently. In terms of performance

394 practice or pedagogy, slow notes are often intended to be connected as smoothly as possible,

395 in order to create a sense of continuity. In terms of genre and harmony, while jazz/blues

396 (on the third dimension) is the most extreme example of this, many genres have harmonies

397 associated with them. For example, the classical genre has fairly structured rules for both

398 harmony and voice leading, but the romantic era relaxed those rules and introduced more

399 complex harmonies. The gradual devolution of those rules and the increase in complexity of

400 harmony continued through the modern and contemporary styles. Although these specific

401 contributions aren’t as strong as some of the others, a glance back at the factor scores plot

402 shows that the older styles: baroque, classical, and romantic, are both negative on the 2nd

403 dimension, as are the simpler harmonies of major and minor. Likewise the newer western

404 styles: impressionist, modern, and contemporary, load positively on the 2nd dimension,

405 along with the more complex harmonies of chromatic, whole tone, and ambiguous.

406 Historically speaking, the whole tone scale gained great popularity with composers in the

407 impressionist era. However, because of the nature of this survey, these results tell us more

408 about the perception of the excerpts rather than the behavior of the participants. Because

409 the excerpts were composed with the intent of varying across all of these musical

410 dimensions, what we see is a sort of validation that there is, in fact, that variety among

411 these excerpts, and that they are different enough to create a large and varied factor space.

### 412 Experiment 2: Musical Adjectives Survey

413

414

415

416

417

418

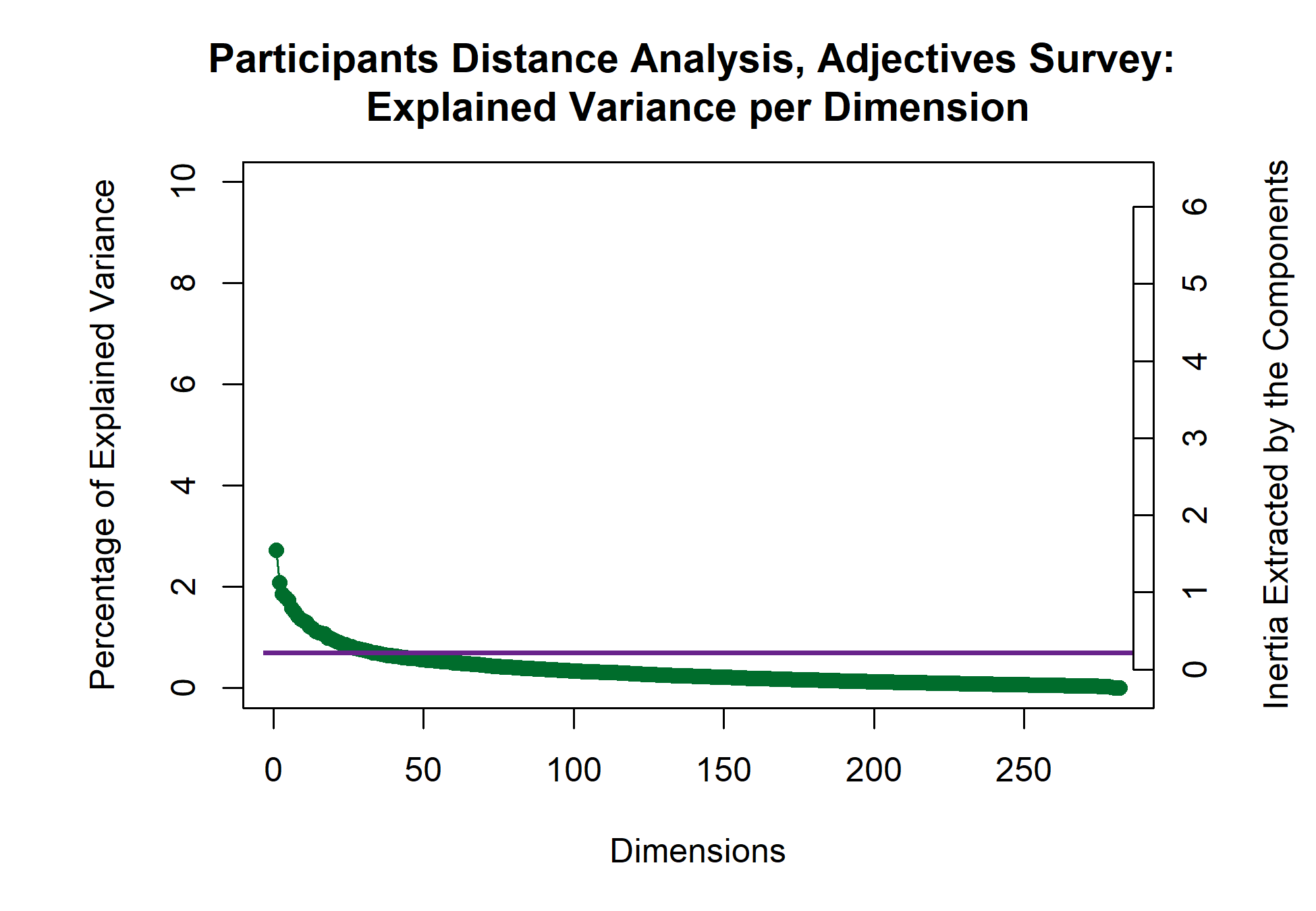
419

420

421

422

**Participants.** The scree plot depicted in Figure [6](#_bookmark4) shows the explained variance per dimension for the distance analysis of participants in the adjectives survey. Again, having a high number of participants means that the dimensionality is high, and each dimension is only extracting a little variance. The first five dimensions all have > 1: 1.66, 1.27, 1.13, 1.09, and 1.06, respectively, but because of



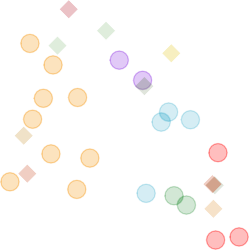
*Figure 6*

423 the high dimensionality here, the first dimension extracts only ~3% of the overall variance.

𝜆

Dimensions 1 and 2

1.0



*.F1*

*dyn.Gra\_Decr*

*genre.Jazz*

*melody.No*

**27**

*e.Mode***1***rn***2**

**7**

*genre.Contemporary*

**6**

*met***1***er.***9***Quad*

*cont.Asce***1***nd***4***ing*

**15**

**8**

*dyn.Loud*

*.F2*

**10**

**9**

**25**

**2**

*motion.Disjunct*

**24 21**

*tic.Legato*

**28**

**18**

**5 22**

*a***2***rti***0***c.Marcat***1***o* **6**

*con harm.M***2***a***9***jor*

*ter.Trip*

*me*

*.Classical*

*genre*

*re.Romantic*

**11**

**3**

*gen*

*artic.Staccato*

*genre.Baroque*

**1**

**2**

**17 30**

*harm.Minor*

*tempo.F3*

*dyn.Soft*

*.F6*

*tempo*

**4**

*t.Undu***2***lat***6***ing*

**13**

*ar*

*tempo.F7*

*tempo.F5*

**3**

*tempo*

*genr*

*harm.Blues*

*/Blues*

*harm.Quin*

*tempo*

0.5

Dimension 2. λ = 0.094. τ = 14%

0.0

−0.5

−1.0

−1 0 1

Dimension 1. λ = 0.123. τ = 18%

*Figure 5*

424 Again, as above, for the purposes of this case study, we’re focusing on the first two

425 dimensions.

426 An MDS analysis of a distance matrix calculated from the pages of the brick revealed

427 significant group differences in how French and American participants described the

428 excerpts, *p*. < .01. The factor scores of the participants are plotted in Figure [7](#_bookmark6), with with

429 group means and bootstrapped confidence intervals shown for those means. The

430 bootstrapping resampling was performed with 1000 iterations. We also analyzed the data

431 using two other participant groupings as factors: gender identity, with three levels: Male,

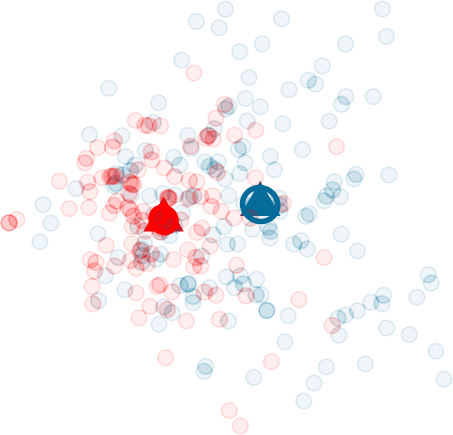
432 Female, or Non-Binary, and level of music training, with three levels: < 2 years, 2-5 years,

433 and >5 years. Neither of these analyses revealed any significant differences between groups.

## Rv Analysis of Participants

Including Group Means and Confidence Intervals

0.2



**AM**

**FR**

0.0

Dimension 2. λ = 1.265. τ = 2%

−0.2

−0.2 0.0 0.2

Dimension 1. λ = 1.66. τ = 3%

*Figure 7*

434 **Excerpts.** The plot in Figure [8](#_bookmark7) shows the explained variance per dimension in the

𝜆

435 analysis of the excerpts contingency table. Although there are no components with

> 1,

436 there are two strong dimensions that extract a majority of the variance. The first two

437 dimensions extract 72.25% of the variance, with the first dimension extracting a majority:

438 50.05%, and the second dimension extracting almost a quarter of the overall variance:

439 50.05%. This plot also suggests that there are multiple ‘elbows,’ at the 3rd, 5th, and 7th

440 dimensions, respectively, with the third and fourth dimensions forming an ‘eigen-plane,’ of

441 two dimensions which extract similar amounts of variance and should be considered

442 together. For this analysis, however, we’re focused on the two first dimensions. Although

443 excerpts 6 and 14 are outliers in the musical qualities survey, for reasons detailed above,

444 they were not outliers in this analysis. We therefore included them in all of the analyses for

445 Experiment 2.

446 The contributions to the first two dimensions are depicted in Figure [9](#_bookmark8). Contributing

447 significantly to the positive end of the first dimension are excerpts from group three (green)

448 and to the negative end are excerpts from group one (yellow). Strong contributions on the

449 positive end of the dimension from the adjectives “Sad,” “Dark,” “Melancholy,” “Slow,”

450 “Mysterious,” “Solemn,” and “Disturbing.” The negative end of the first dimension is

451 defined by the adjectives “Fast,” “Happy,” “Dancing,” “Colorful,” and “Bright.” The

452 second dimension is dominated by excerpts from group 4 (red) in the positive direction and

453 group 2 (blue) in the negative direction. Two excerpts from group 3 also contribute

454 significantly, excerpts 7 in the positive direction and excerpt 10 in the negative direction.

455 The columns contributing strongly in the positive direction are “Aggressive,” “Fast,”

456 “Disturbing,” “Mysterious,” “Surprising” and “Complex.” The columns contributing in the

457 negative direction are “Warm,”Soft“,”Happy“,”Slow“,”Round“, and”Light”.

458 The barplots in Figure [10](#_bookmark9) show the bootstrap ratios calculated for the rows and

459 columns. Here we’ve included all of the rows and columns, because it’s useful to see both

30

0.2

# Explained Variance per Dimension

40

50

60

0.3

0.4

0 5 10 15 20 25 30



Percentage of Explained Variance

0

10

20

0.0

0.1

Inertia Extracted by the Components

Dimensions

*Figure 8*

460 which are significant and which are not. This is an inferential method that tells us is how

461 consistently each of the observations and variables load on the first two dimensions. The

462 threshold in this case is *p* < .05. From this we get an idea of which of the rows and

463 columns are stable, in other words, which ones tended to be rated in a certain way

464 consistently across all participants, and also how likely these are to be observations

465 reflective of the population as a whole. In this plot, the more extreme value of the

466 bootstrap ratio, the more likely that it is a reflection of the ‘real’ value. The values in the

467 center of each plot that are grayed out identify the rows or columns that are not

468 consistently loading on the dimensions. With the observations and variables ordered like

469 this, it makes it easy to see how the consistently the clusters are distributed in the space.

470 This plot was not included for experiment 1 because it would be less informative given

0.10

0.05

Contributions

0.00

Excerpt.3

−0.05

−0.10

## Component 1

Rows

Excerpt.4

Excerpt.13

*Contributions*

Columns

0.10

Excerpt.19

Excerpt.23

Excerpt.26

Excerpt.29

Contributions

Bright

Colorful Dancing

Fast

Happy

0.05

0.00

Excerpt.18

Excerpt.24

Excerpt.27

Dark

Disturbing

Melancholy Mysterious

Sad

Slow Solemn

−0.05

−0.10

## Component 2

Excerpt.5 Excerpt.7

Excerpt.10

Rows

Excerpt.10

Excerpt.11

Excerpt.20

Columns

0.1

Contributions

0.0

0.1

0.0



Excerpt.1 Excerpt.6

Excerpt.7

Excerpt.16

Excerpt.25

Excerpt.28

Excerpt.29

Contributions

Aggressive Complex Disturbing

Fast

Happy

Light

Mysterious

Round Slow

Soft

Surprising

Warm

−0.1

−0.1

*Figure 9*

471 what the survey in experiment 1 was assessing. Experiment 1 doesn’t evaluate the behavior

472 of participants, but the nature of the excerpts. Note that there are far more significant

473 bootstrap ratios than there are significant contributions. That just means that while not

474 everything is contributing, overall the model seems to be stable. Fewer significant

475 bootstrap ratios would suggest that there was a greater amount of variance in the

476 observations and variables than were accounted for, at least in the first two dimensions.

477 Looking at the nonsignificant values for the adjectives may inform our understanding of the

478 participants’ use of the adjectives. ‘Incisive,’ ‘transparent,’ ‘poweful,’ ‘dense,’ ‘round,’ and

479 ‘sparse,’ are all nonsignificant on the first dimension, and ‘weak,’ ‘dull,’ ‘sparse,’ ‘valiant,’

480 and ‘short’ are all nonsignificant on the second dimension. All but ‘sparse’ are significant

481 on one dimension or the other. Looking at the column sum for ‘sparse’ tells us that it was

482 used, so this isn’t an effect of participants not using this word. It’s more likely that ‘sparse’

483 doesn’t really fit into the Valence-arousal plane. It’s a neutrally valenced word that could

484 describe excerpts that fall anywhere within that plane. ‘Weak’ and ‘transparent’ give us

485 another important perspective. These were the two least commonly used adjectives, but

486 the fact that they are consistently loading on one dimension or the other suggests that

487 when they were used, they were used in the same way.

Component 1

Rows

*Bootstrap ratios*

Columns

20 30

Bootstrap ratios

Excerpt.2 Excerpt.2 Excerpt.1 Excerpt.4 Excerpt.1 Excerpt.2 Excerpt.9 Excerpt.1 Excerpt.2 Excerpt.1 Excerpt.8 Excerpt.1 Excerpt.2 Excerpt.2 Excerpt.2 Excerpt.2 Excerpt.6

Bootstrap ratios

Fast Happy Dancing Colorful Bright Light Valiant Warm Surprising

Aggressive Strong Exotic Varied Short Complex Incisive Transparen

0 0

Excerpt.11 Excerpt.12 Excerpt.30 Excerpt.17 Excerpt.15 Excerpt.10 Excerpt.21 Excerpt.5 Excerpt.18 Excerpt.24 Excerpt.7 Excerpt.3 Excerpt.27

Powerful Dense Round Sparse Monotonous

Weak Soft Long Dull

Disturbing Solemn Mysterious Melancholy

Dark Slow Sad

−20

−40

20

Bootstrap ratios

10

0

−10

−20

Component 2

Rows

Excerpt.11 Excerpt.28 Excerpt.29 Excerpt.20 Excerpt.30 Excerpt.10 Excerpt.15 Excerpt.8 Excerpt.23 Excerpt.21 Excerpt.3 Excerpt.19 Excerpt.17 Excerpt.13 Excerpt.26 Excerpt.27

−30

−60

20

Bootstrap ratios

10

0

Excerpt.14 Excerpt.4 Excerpt.22 Excerpt.24 Excerpt.9 Excerpt.18 Excerpt.12 Excerpt.5 Excerpt.16 Excerpt.2 Excerpt.7 Excerpt.25 Excerpt.1 Excerpt.6

−10

−20

Columns

Warm Soft Happy Slow Light

Melancholy Round Solemn Sad Dancing Bright Colorful

Transparent Long

Weak Dull Sparse Valiant

*Figure 10*

Short Monotonous

Strong Dense Powerful Exotic Incisive Varied Dark Complex Surprising Mysterious Disturbing

Fast Aggressive

488 **Discussion.** The factor maps below show the row and column factor scores for the

489 American and French participants. These are once again symmetric plots, interpretation is

490 the same as the factor plot for the musical qualities. There’s a clear valence-arousal plane

491 apparent for both, and in both cases valence seems to define the first dimension and

492 arousal defines the second dimension. However, the difference in the amount of variance

493 extracted by the first two dimensions between the French and American participants is

494 notable. The French data show a weaker first dimension but a stronger second dimension

495 relative to the Americans, both in terms of variance extracted (tau), effect size (lambda).

496 This tells us that French participants were less affected by the excerpts than the American

497 participants, but they responded more to the arousal of the excerpts. There are also

498 differences in how the adjectives and the excerpts are distributed in the space. One clear

499 example is that Excerpt 6 is in quadrant two in the American plot, but quadrant one in the

500 French. This is a small change, but it suggests that the French participants were more

501 likely to assign negative valence to this excerpt, and American Participants were more

502 likely to assign positive valence. For the adjectives, ‘bright’ and ‘dancing’ are directly on

503 top of one another in the American plot, but there is some space between the two in the

504 French plot. It’s possible that this reflects the idea that although the meaning is shared

505 between languages, there are semantic or associational differences between the words.

506 Additionally, a post-hoc Multiple Factor Analysis revealed the following in terms of

507 the semantic and perceptual differences between French and American participants.

Contributions to the Excerpts Factor Scores



**FR**

**AM**

**FR**

**cerpt.1**

**AM Excerpt.25**

**AM**

**FR Excerpt.2**

**EFRxcerpt.16**

**FR**

**Excerpt.9**

**AM**

**FR**

**t.18**

**FR**

**FR**

**FR**

**Excerpt.24**

**pt.4**

**Excerp**

**AM**

**AtM.27**

**FR**

**.28**

**Excerpt**

**FR**

**t.11**

**AM**

**Excerp**

**FR**

**AM**

**Excerpt.10**

**FR**

**30FR**

**AM**

**Excerpt.**

**erpt.20**

**EAMxc**

**Excerpt.8 cerpt.29**

**FR FR**

**Ex**

**AM**

**AM**

**FR**

**FR AM**

**Excerpt.3**

**Mcerpt.15 FR AM FR**

**cerpt.17**

**EAx**

**AM FR**

**Excerpt.21**

**AM**

**Ex**

**AMpt.14**

**AM**

**t.19**

**Exce**

**FeRArMpt.26**

**AM**

**AM**

**A.1M 3FRExcerp FR**

**cerpt.23**

**Exc**

**Excerpt**

**Ex**

**Excerp**

**AM**

**AM**

**Excerpt.22**

**FR**

**FR**

**AM**

**FR**

**Excer**

**Excerpt.7**

**AM**

**FR**

**erptA.M5**

**FR**

**AM**

**Exc**

**AM**

**FR**

**Excerpt.6 Excerpt.12**

**Ex**

0.2

0.3

Contributions to the Adjectives Factor Scores

0.2

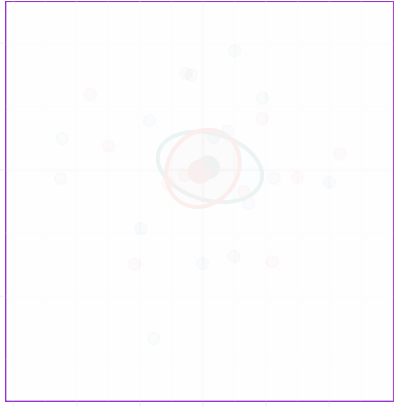
Excerpt.3

Contributions

⊤

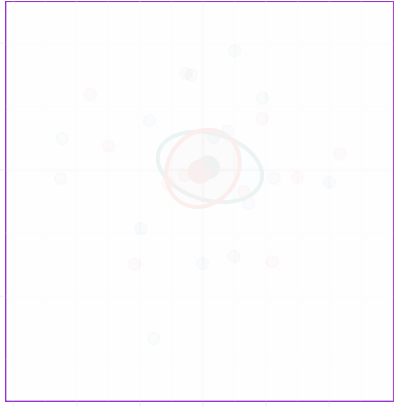
Dimension 2. λ = 7.52. τ = 6%

Dimension 2. λ = 7.52. τ = 6%



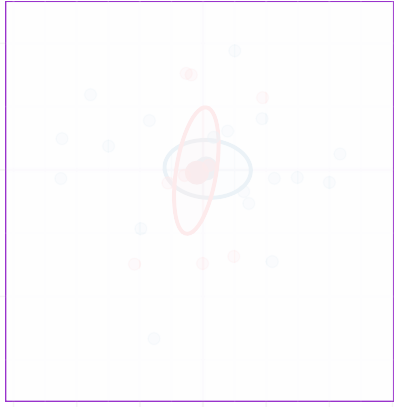
𝜆

to focus on the first two dimensions, with 𝜆 𝜆



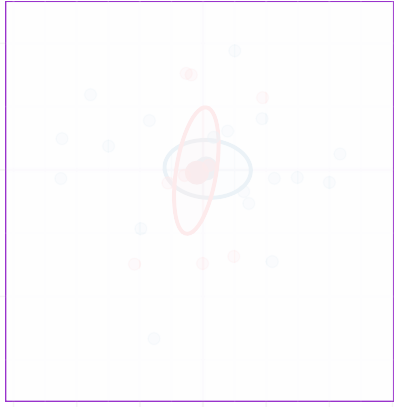
**F**

**M**



to focus on the first two dimensions, with 𝜆 𝜆

𝜆



**AM**

**FR**

to focus on the first two dimensions, with 𝜆 𝜆

𝜆

**AM**

**Complex AAM ggr**

**risinAgM**

**AMFR FR**

**ieAMd FIRncisive**

**FR**

**ong AM Powerful**

**AM FRAM AM**

**nseAM Weak FR**

**AM Short FR AMFR**

**FR AM**

**ColoArMful**

**FR**

**Sparse**

**AM AM**

**L**

**AM FR**

**AAMM**

**AM**

**AM**

**t**

**AM**

**Sad**

**FR**

**FR**

**FR Solemn Melanc**

**AM**

**und**

**FR**

**Slow**

**AM**

**FR**

**FR**

**FR**

**WarmAM**

**AM**

**FSRoft**

**M**

**FR**

**Ro**

**Light**

**DancFRing**

**HFaRppAy**

**Brigh**

**holy**

**ong**

**AM**

**Valiant**

**FR**

**Dull**

**AM FR**

**FR**

**AM**

**Exotic De**

**FR**

**FR**

**Str**

**DaArMk**

**FR**

**AM**

**Surp**

**Var**

**urbinAMg**

**Dist**

**essive**

**FR**

**s**

**FRMysteriou**

**FR FR**

**ast**

**F**

**FR**

0.1

Dimension 2. λ = 0.02. τ = 12%

Dimension 2. λ = 0.015. τ = 11%

0.0

0.0

−0.1

−0.2

−0.2

−0.3

−0.25

0.00

Dimension 1. λ = 0.029. τ = 16%

0.25

−0.2 0.0 0.2

Dimension 1. λ = 0.033. τ = 15%

*Figure 11*

### 508 Experiment 3: Combined Surveys

509 Experiment 3 used the pseudo-contingency tables from experiments 1 and 2 together.

510 Since excerpts 6 and 14 were excluded from analysis for experiment 1, we also removed

511 those rows from the contingency table for experiment 2. This is so that the dimensions of

512 the two tables for this PLSC would be conformable (remember that we need the same rows

513 or columns in both tables for this analysis). The point of this experiment is to identify the

514 strongest covariance between the two tables - that is, the strongest shared signal between

515 two data tables. Now, this is not to say that these two tables are evaluating the same thing.

516 Instead it allows us to see what is most common between two sets of different information -

517 how often an excerpt was associated with *both* a musical quality and an adjective. The

518 visualizations below allow us to see which variables from each of the two tables correspond

519 with one another; which adjectives are associated with which musical dimensions. Even

520 though both individual tables have their own factor spaces, plotting the common factor

521 space between the two should allow us to see which excerpts are separated from one

522 another using data from both surveys.

523 **Results.** This analysis revealed two dimensions that extracted the majority of the

524 variance (83.60%). Of that total extracted by the first two dimensions, the first dimension

525 extracted 64.35% and the second dimension extracted 19.26%. The scree plot below shows

526 that it’s possible that there are two elbows in this graph, at the 3rd and 5th dimensions.

527 The 3rd and 4th dimensions are also significant, extracting 6.02% and 3.67% of the

528 variance, respectively. Interpretations of the third dimension and beyond is beyond the

529 scope of this paper, but seeing that there are multiple significant dimensions beyond the

530 second does provide a possible future direction using this method.

531 The plot below shows which variables from each data table load the most on the first

532 and second dimensions. For the purposes of this visualization, we are showing only the

533 variables for which 70% or more of the variance is explained. The nature of the PLSC also

534 suggests that these are the variables that are most associated with one another between the

535 two tables. The strongest signal on the first dimension juxtaposes the slow and legato

536 musical qualities in the positive direction with the fast, staccato, marcato, and conjunct

537 musical qualities in the negative direction. The adjectives associated with the qualities in

538 the positive direction are “Dark,” “Dull,” “Long,” “Melancholy,” “Sad,” “Slow,” “Solemn,”

539 and “Weak.” The adjectives associated with the negative direction are “Bright,” “Colorful,”

540 “Dancing,” “Fast,” “Happy,” and “Light.”

60

# PLSC Music Features: Inertia Scree Plot

40

50

60

70

80 100

0 5 10 15 20 25



Percentage of Explained Variance

0

10

20

30

0

20

40

Inertia Extracted by the Components

Dimensions

*Figure 12*

541 The second dimension identified in the positive direction major harmony and mezzo

542 dynamics, associated with “Light,” “Round,” “Soft,” and “Warm.” The negative direction

543 is driven by the impressionist genre being associated with “Aggressive,” “Complex,”

544 “Dense,” “Disturbing,” “Powerful,” and “Surprising.”

545 Contributions and loadings are similar, but not exactly the same. Here were see that

546 there are quite a few more variables that contribute significantly to these dimensions than

547 for which a significant portion of the variance is explained. We do see similar groups,

548 however: on the first dimension, the tempo variables are contributing significantly, along

549 with some from harmony, density, genre, dynamics, motion, range, and articulation. The

550 adjectives contributing significantly are Bright, colorful, Dancing, Fast, Happy, Light, and

551 Valiant in the negative direction, and Dark, Dull, Long, Melancholy, Monotonous, Sad,

Loadings for variables for factor plot 1 Loadings for variables for factor plot 2

0.4

0.2

tempo.F7 motion.Conjunct artic.Marcato

artic.Staccato

genre.Impressionist

0.2

tempo.F6

0.0

Signed Loadings

tempo.F3

tempo.F2

artic.Legato

Bright Colorful

Dancing

Dark

Dull

Fast

Happy Light

Long Melancholy

Sad Slow Solemn

Weak

Signed Loadings

Aggressive

Complex

Dense

Disturbing

Surprising

Varied

0.0

harm.Major

Light

Round

Soft

Warm

−0.2 −0.2

*Figure 13*

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

Slow, Solemn, and Weak in the positive direction. What’s notable here is that while some of these variables did contribute significantly in the plots above (see Figure **??** and Figure [5](#_bookmark5)), some didn’t contribute much at all and fell near the barycenter of the factor plot. We also see that this juxtaposes some negatively and positively valenced adjectives, which allows us to identify which of the musical qualities contributes to the valence dimension. The second dimension tells us a similar story. Here we see more of the harmony variables, along with one tempo variable, some density, genre, a few dynamics, contour, motion, range, and articulation. The adjectives contributing negatively are Aggressive, Complex, Dense, Disturbing, Incisive, Mysterious, Powerful, Surprising, and Varied, and those contributing positively are Light, Round, Soft, Transparent, and Warm. Again we see similar effects of variables that may not have contributed significantly to their respective plots above, but are contributing significantly here. Also, this second latent variable seems to be defining the arousal dimension.

**Discussion.** The factor score plots for this analysis shows that the first two sets of

latent variables extracted by the analysis effectively separate the groups of excerpts into the clusters defined in the HCA for the adjectives survey. This factor plot shows us how the strongest correlated signal between the two data tables separates Excerpts groups 2 and 3, but groups 1 and 2 didn’t contribute much to this dimension, instead contributing

Contributions to the First Latent Variables

0.05

Signed Contributions

harm.Major

tempo.F6

tempo.F7

density.D4

genre.Classical

dyn.Loud

cont.Undulating

motion.Conjunct range.R.Vwi artic.Marcato

artic.Staccato

0.00

harm.Ambiguous harm.Whole tone

tempo.F3

tempo.F2

tempo.F1

density.D2

dyn.Soft

motion.Disjunct

artic.Legato

Bright

Colorful Dancing

Dark

Dull

Fast Happy

Light

Long Melancholy Monotonous

Sad Slow

Solemn

Valiant

Weak

−0.05

genre.Impressionist

−0.10

0.15

Contributions to the Second Latent Variables

0.10



Signed Contributions

harm.Ambiguous harm.Chromatic harm.Quin

tempo.F6

density.D6

genre.Contemporary

genre.Impressionist

dyn.BothS&L

dyn.Gra\_Cresc

cont.Ascending

cont.Monotone motion.Disjunct

0.05

harm.Minor

harm.Modal

artic.Marcato

Aggressive Complex Dense Disturbing

Incisive

Mysterious

Powerful

Surprising

Varied

0.00

harm.Major

density.D3

genre.Romantic

dyn.Mezzo

dyn.Soft

cont.Arch

motion.Conjunct

range.R.Mod

artic.Legato

Light

Round

Soft

Transparent

Warm

−0.05

−0.10

*Figure 14*

570 to the 2nd latent variables. The second latent variable separates Groups 1 and 4, with

571 Groups 2 and 3 more barycentric. This suggests that, generally speaking, the excerpts that

572 were clustered in groups 2 and 3 are those that could be defined by positive and negative

573 valence, respectively, and those in groups 1 and 4 would be defined more by high and low

574 arousal. That being said, these excerpts are not defined *exclusively* along these dimensions,

575 but rather more by one than the other. For example, excerpt 26 is characterized by being

576 one of the most extreme examples of positive valence, but doesn’t score as highly on the

577 arousal dimension, similarly with excerpt 27 with negative valence. This is contrasted with

578 excerpt 7, which is one of the most negatively valenced stimuli, but also scores very high on

579 arousal, although the barycenter for that group is near the origin of that plot.

Latent Variable Plot 1 Latent Variable Plot 2

1.0 1.0

Warm

Surprising

Round Slow

Soft

Mysterious

Happy

Light

Aggressive Complex Disturbing

Fast

Contributions

Excerpt.28

Excerpt.29

Excerpt.25

Excerpt.16

Excerpt.1 Excerpt.6

Excerpt.7



Excerpt.11 Excerpt.28 Excerpt.29 Excerpt.20 Excerpt.30 Excerpt.10 Excerpt.15 Excerpt.8 Excerpt.23 Excerpt.21 Excerpt.3 Excerpt.19 Excerpt.17 Excerpt.13 Excerpt.26 Excerpt.27

Bootstrap ratios

Powerful Dense Round Sparse Monotonous

Weak Soft Long Dull

Disturbing Solemn Mysterious Melancholy

Dark Slow Sad

Excerpt.11 Excerpt.12 Excerpt.30 Excerpt.17 Excerpt.15 Excerpt.10 Excerpt.21 Excerpt.5 Excerpt.18 Excerpt.24 Excerpt.7 Excerpt.3 Excerpt.27

Fast Happy Dancing Colorful Bright Light Valiant Warm Surprising

Aggressive Strong Exotic Varied Short Complex Incisive Transparen

Bootstrap ratios

Excerpt.2 Excerpt.2 Excerpt.1 Excerpt.4 Excerpt.1 Excerpt.2 Excerpt.9 Excerpt.1 Excerpt.2 Excerpt.1 Excerpt.8 Excerpt.1 Excerpt.2 Excerpt.2 Excerpt.2 Excerpt.2 Excerpt.6

Bootstrap ratios

0.5 0.5

0.0 0.0

Ly 1

Ly 2

−0.5 −0.5

−1.0 −1.0

−1.5 −1.0 −0.5 0.0 0.5 1.0

Lx 1

−1.0 −0.5 0.0 0.5 1.0

Lx 2

*Figure 15*

### 580 General Discussion

581 Although this study was designed to evaluate the sensory or cognitive response to

582 music, and not specifically the emotional response, there is significant overlap in the results

583 observed here and the results of the work investigating music and emotion. The

584 appearance of the valence-arousal plane in the results of experiment 2 was not unexpected,

585 even though the adjectives we selected were not intended to be explicitly emotional. This

586 goes to show diﬀicult it is to avoid any emotional content when selecting descriptors, and

587 from another perspective, how much emotional contagion the musical examples carry.

588 Overall, this supports the idea that the first two dimensions on which music is judged

589 holistically are valence and arousal. Some of the results discussed in Experiment 1 require

590 more explanation. In experiment 1, there was an issue of having two individual excerpts

591 dominate the factor space, numbers 6 and 14, which did not happen in experiment 2. One

592 of the ways in which CA is different from PCA is that PCA is usually unweighted. CA, on

593 the other hand, makes use of weights and masses to find the average observation.

594 Information that is common, therefore, falls towards the center of the plot, while

595 information that is further from the average, in other words, more rare, ends up further

596 from the center of the factor plots. [cite] Therefore, if a survey like the one used in

597 experiment 1 includes a item that is wildly different than the others in the set, the ratings

598 will be very different, and that item will dominate the factor space. In this case we have

599 two such examples: excerpts 6 and 14. Excerpt 6 was written as a Steve-Reich-esque

600 minimalist, ostinato based excerpt, and excerpt 14 was written to be jazzy. The reason this 601 effect occurs with the first survey and not the second is that the musical qualities on which 602 the excerpts were rated were explicit and designed to separate the excerpts along the

603 various musical dimensions, while the adjectives survey was designed to evaluate the

604 excerpts more generally on holistic qualities. Excerpt 6 still appears as a minor outlier in 605 the visualizations for the second survey, but does not dominate the space the way it does in 606 the results of the first. What we did to mitigate that is to use those two excerpts as

607 *supplementary projections*, sometimes also referred to as *out of sample observations*. This 608 allows us to evaluate what information is shared by those outliers with the other elements 609 in the dataset without having them dominate the visualization of the factor space. If, when 610 we projected those values into the factor space, they projected onto the origin or very close 611 to it, we would know that those observations shared no information with the other

612 variables. The fact that they are where they are offers support to the idea that the first 613 survey separates the excerpts approximately by genre. Because the ‘genre’ information 614 isn’t shared with the other observations, they are being projected onto the space sharing

615 only the information that does not deal with genre, like tempo or range. What this tells us

616 is that musical qualities surveys captured a result that may have characterized by 4-6

617 factors, each approximating genre and the qualities associated with that genre and the 618 general affective space captured an entirely different set of information about the stimuli 619 and the perception of the stimuli.

620 The hierarchical cluster analyses revealed different groupings in how the stimuli were

621 rated between the two surveys. The PLSC then showed that when including both sets of

622 data, there was a coherent interpretable factor space on which the excerpts were plotted. 623 There are a number of ways to further disambiguate the results of the surveys. One way 624 would be to run a MFA, similar to the one above that plotted the difference between

625 French and American raters on the adjective survey. This would allow for a number of

626 different interpretations. Firstly, it would calculate the overall factor space for the excerpts,

627 including all of the data from both surveys, without separating out the first and second

628 dimensions to plot them separately. It would also identify the specific partial factor scores 629 for each of the data tables within that factor space that would allow for the interpretation 630 of the relative differences between the data tables. The drawback to both of these, however, 631 is that unlike the separate correspondence analyses we ran above, where the row and

632 column scores can be plotted in the same space, neither MFA nor the PLSC allow for that 633 type of visualization. That being said, because different types of analysis reveal different 634 aspects of the data, running both analyses can provide a broader understanding of the

635 data, and each could provide explanations for what remains ambiguous in the other. An 636 important overall takeaway from this is that with a deep general understanding of the 637 stimuli, we may be able to predict the approximate dimensionality of the solution factor 638 space. In the first survey, the solution was that the stimuli were largely separated along

639 genre or stylistic lines. One issue that arose with this is that there was only one example of 640 minimalist and jazz music. To have a solution in which we didn’t see these specific excerpts 641 as outliers, but as coherent members of a factor space, we would need more examples of

642 those styles. This suggests that when creating surveys or designing stimuli, we should keep 643 in mind that we need multiple items per group, or presumed dimension. This is not to say 644 that we will always be able to a priori predict the factor space of the solution. For example, 645 experiment 2 may also have benefitted from more minimalist or jazz examples - in a system 646 in which the overall structure is obtained by evaluating the stimuli holistically, having a

647 single outlier will necessarily distort the space. Either because it is an outlier in sensory

648 terms or because it is the only stimulus against which there is no direct reference. This in a

649 way embodies the issue described in the introduction, where we have a single dimension

650 that is noisy. This really only applies to experiment 2. The noise comes from the fact that

651 participants were likely to be less familiar with mimalism and/or jazz than the trained 652 musicians who took the QS, but the reason the results are overall robust to that noise is 653 that the participants were not asked to rate the excerpts on any explicit dimensions or 654 qualities.

### 655 Limitations & future directions

656 Although we evaluate the scores and ratings of participants from different countries, 657 we recognize that the issue of multiculturality is not addressed to a significant degree in 658 this study. The sample was still largely students, and France and the United States share 659 similar musical cultures. To truly address this question, it would be very interesting to

660 include participants from multiple, contrasting musical cultures, with languages that are

661 more distinct than English and French. This presents new problems, however, as the

662 specific musical qualities included in the surveys may not all apply to or translate well to

663 other musical cultures. Harmony, for example, is a concept that is developed to a

664 significant degree in western music, but melody or rhythm may be the fundamental focus 665 of other musical cultures (cite patel here? I forget.). Another question that fell beyond the 666 scope of this study is the concept of semantic drift between languages. Although illustrated 667 in Figure [11](#_bookmark10), the source of the differences between French and American participants is not 668 entirely clear. We humbly hazard to guess that some of the sources of the difference include 669 aspects of perception that extend beyond the musical. These could be linguistic sources,

670 such as the physical characteristics of the words themselves, the cultural associations with 671 the words, or the frequency of use in either language. Diving more into those questions of 672 linguistics and semantic drift between languages would be a fascinating future study.

673 Another interesting study would be to repeat this study using adjectives from specific

674 domains or that that avoid explicit emotional or musical content, to see how music maps

675 onto different sensory spaces. For example, ‘moist,’ ‘slimy,’ ‘dry,’ ‘puckered,’ ‘smooth.’

676 Although some of these adjectives may carry musical weight, in the context of other words 677 that all relate to haptic sensation, it may provide some interesting feedback regarding how 678 the music maps into other sensory domains. Finally, using these studies may provide pilot 679 work for the way in which people without language react to music, nonverbal autistic

680 people, for example. Whereas this study explicitly uses language as an interlocutor for

681 music perception, it offers insight into ways to better communicate with people who do not

682 have that ability.

683 **Conclusions**

684 Expanding the collection and analytical paradigms, and thus expanding scientific

685 scope and perspecive, has the added benefit of increasing reach. By expanding the ways in 686 which we collect data, we are able to more readily and consistently reach participants who 687 might normally be excluded from everday research paradigms, specifically racially and

688 ethnically diverse populations, poorer populations, those with limited access to

689 transportation, or who have a disability, or are immunocompromised. By developing

690 investigative paradigms that are accessible on mobile platforms and that reduce participant

691 demand while maintaining rigor and integrity, we are likely to be able to reach a much 692 greater subset of the population. If we are able to pair this kind of data gathering with 693 appropriate analysis, we can maintain the standards of scientific integrity that we as a

694 community expect with traditional hypothesis testing. The literature to date in the music 695 cognition domain has focused on a fairly small subset of the multivariate analyses available 696 to investigate these questions. As presented here, the number of ways that exist to analyze 697 the data from a single set of experiments is considerable, and the results of each analysis 698 illuminate different parts of the story the data are telling. Not every form of analysis is

699 appropriate in every context, but understanding how, and perhaps more importantly when,

700 to apply a technique or type of analysis is an important to uncovering new perspectives or

701 insights.

702 **References**

703 Abdi, H., & Williams, L. J. (2010). *Correspondence Analysis* (N. Salkind, Ed.).

704 Sage.

705 Abdi, H., & Williams, L. J. (2013). *Partial Least Squares Methods: Partial Least*

706 *Squares Correlation and Partial Least Square Regression: Vol. II* (B. Reisfeld &

707 A. N. Mayeno, Eds.; pp. 1453–1454). Springer Science+Business Media, LLC.

708 <https://doi.org/10.1007/978-1-62703-059-5>

709 Abdi, H., Williams, L. J., & Valentin, D. (2013). Multiple factor analysis: Principal

710 component analysis for multitable and multiblock data sets. *Wiley*

711 *Interdisciplinary Reviews: Computational Statistics*, *5*(2), 149–179.

712 <https://doi.org/10.1002/wics.1246>

713 Ares, G., Deliza, R., Barreiro, C., Giménez, A., & Gámbaro, A. (2010). Comparison 714 of two sensory profiling techniques based on consumer perception. *Food Quality* 715 *and Preference*, *21*(4), 417–426. <https://doi.org/10.1016/j.foodqual.2009.10.006>

716 Benzécri, J.-P. (1973). *L’analyse des données.* (p. 615). Dunod.

717 Berry, K. J., Johnston, J. E., & Mielke, P. W. (2011). Permutation methods. *Wiley*

718 *Interdisciplinary Reviews: Computational Statistics*, *3*(6), 527–542.

719 <https://doi.org/10.1002/wics.177>

720 Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005).

721 Multidimensional scaling of emotional responses to music: The effect of musical

722 expertise and of the duration of the excerpts. *Cognition and Emotion*, *19*(8),

723 1113–1139. <https://doi.org/10.1080/02699930500204250>

724 Borg, I., & Groenen, P. J. F. (2005). *Modern Multidimensional Scaling* (2nd ed.,

725 Vol. 36, pp. 1–614). Springer Science+Business Media, Inc.

726 Bruner II, G. C. (1990). Music, Mood, and Marketing. *Journal of Marketing*,

727 *October*, 94–104.

728 Coombs, C. H., Milholland, J. E., & Womer, F. B. (1956). The assessment of partial

729 knowledge. *Educational and Psychological Measurement*, *16*(1), 13–37.

730 <https://doi.org/10.1177/001316445601600102>

731 Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of 732 emotion bridged by continuous gradients. *Proceedings of the National Academy* 733 *of Sciences of the United States of America*, *114*(38), E7900–E7909.

734 <https://doi.org/10.1073/pnas.1702247114>

735 Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis* (pp.

736 1–376). Academic Press.

737 Hesterberg, T. (2011). Bootstrap. *Wiley Interdisciplinary Reviews: Computational*

738 *Statistics*, *3*(6), 497–526. <https://doi.org/10.1002/wics.182>

739 Juslin, P. N., & Sloboda, J. A. (Eds.). (2010). *Handbook of music and emotion:*

740 *Theory, research, applications.* (pp. xiv, 975–xiv, 975). Oxford University Press.

741 Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students.

742 *Journal of Abnormal and Social Psychology*, *28*(3), 280–290.

743 <https://doi.org/10.1037/h0074049>

744 Kopacz, M. (2005). Personality and music preferences: The influence of personality

745 traits on preferences regarding musical elements. *Journal of Music Therapy*,

746 *42*(3), 216–239. <https://doi.org/10.1093/jmt/42.3.216>

747 Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least 748 Squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage*, 749 *56*(2), 455–475. <https://doi.org/10.1016/j.neuroimage.2010.07.034>

750 Madsen, C. K. (1997). Emotional Response to Music as Measured by the

751 Two-Dimensional CRDI. *Journal of Music Therapy*, *34*(3), 187–199.

752 <https://doi.org/10.1093/jmt/34.3.187>

753 Meyners, M., & Castura, J. (2014). Check-All-That-Apply Questions. In *Novel* 754 *techniques in sensory characterization and consumer profiling* (pp. 271–306). 755 CRC Press/Taylor & Francis. <https://doi.org/10.1201/b16853-12>

756 Osgood, C. E., & Suci, G. J. (1955). Factor analysis of meaning. *Journal of*

757 *Experimental Psychology*, *50*(5), 325–338. <https://doi.org/10.1037/h0043965>

758 Rodà, A., Canazza, S., & De Poli, G. (2014). Clustering affective qualities of 759 classical music: Beyond the valence-arousal plane. *IEEE Transactions on* 760 *Affective Computing*, *5*(4), 364–376.

761 <https://doi.org/10.1109/TAFFC.2014.2343222>

762 Wallmark, Z. (2019). A corpus analysis of timbre semantics in orchestration

763 treatises. *Psychology of Music*, *47* (4), 585–605.

764 <https://doi.org/10.1177/0305735618768102>

765 Wedin, L. (1969). Dimension Analysis of Emotional Expression in Music. *Swedish*

766 *Journal of Musicology*, *51*, 119–140.

767 Wedin, L. (1972). Evaluation of a Three-Dimensional Model of Emotional

768 Expression in Music. *The Psychological Laboratories*, *54*(349), 1–17.