Cognitive Music Listening Space: A Multivariate Approach

Brendon Mizener[1], Mathilde Vandenberghe[2], Hervé Abdi[1], & Sylvie Chollet[2]

[1] University of Texas at Dallas

[2] Junia, Univ. Artois, Université de Liège, Univ. Littoral Côte d'Opale, UMRT 1158

BioEcoAgro, F-62000 Arras, France

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Brendon Mizener, 800 W. Campbell Rd., Richardson Tex. E-mail: bmizener@utdallas.edu

Abstract

French and American participants listened to new music stimuli and evaluated the stimuli using either adjectives or quantitative musical dimensions. Results were analyzed using Correspondence Analysis (CA), Hierarchical Cluster Analysis (HCA), Multiple Factor Analysis (MFA), and Partial Least Squares Correlation (PLSC). All except the HCA used Bootstrapping and Permutation testing for inferences. French and American listeners differed when they described the musical stimuli using adjectives, but not when using the quantitative dimensions. The present work serves as a case study in research methodology that allows for a balance between relaxing experimental control and maintaining statistical rigor.

Keywords: Music, Perception, Cognition, Multivariate Analyses

Word count: 5631

Cognitive Music Listening Space: A Multivariate Approach

We have a data collection problem: World events over the last year have shown that we need to be able to collect good data outside of the lab. In the lab, because we control error sources, we measure, on relatively small sets of observations, a few well-defined, quantitative variables, analyzed using standard techniques such as analysis of variance (ANOVA). But, with the labs closed (remember COVID?), how can we collect good data? Away from the controlled environment of the lab, quantitative variables are hard to measure, but we can collect, on large sets of observations, qualitative variables that can only be analyzed by newer multivariate techniques. In the present paper, we present a case study illustrating this tradeoff.

"Doesn't beer taste better in a bar? Or when you're listening to your favorite song?" The present study was designed to quantify a music listening 'space' that captures objective stimulus and cognitive dimensions for the sake of investigating cross-modal sensory mapping between beer drinking and music listening. It also addresses other questions: Are there differences in how people from different countries — and by extension musical cultures — perceive and describe music? What parallels exist between stimulus dimensions and cognitive dimensions of music?

For the present study, we have defined stimulus dimensions as quantitative musical qualities, such as tempo, range, and meter and cognitive dimensions as qualitative descriptions of music, such as "dark," "warm," and "round." These cognitive/qualitative dimensions are similar to the commonly investigated affective or emotional dimensions, but do not specifically assess affective quality. We designed three experiments to quantify individual and combined spaces for these concepts, using separate surveys. The first experiment included highly trained musicians and featured a multiple choice survey about the stimulus dimensions; the second included participants with any level of music training performing a check-all-that-apply task (CATA, Katz & Braly, 1933; Meyners & Castura,

2014; Coombs et al., 1956); the third experiment incorporated both surveys in a single analysis.

To analyze our data, we selected a set of multivariate analyses that allowed us to visualize answers to each of our questions. The mental spaces revealed by the individual surveys were calculated and visualized using Correspondence Analysis (CA), a method similar to Principal Components Analysis (PCA) that analyzes qualitative data. We used Multidimensional Scaling (MDS), a distance analysis, to visualize the differences between participants and participant groups. To find parallels between the surveys, we used Partial Least Squares Correlation (PLSC), a method that analyzes two data tables with different sets of variables measured on the same observations. We used a Multiple Factor Analysis to evaluate how French and American participants' responses differed. Each of these analyses provide different visualizations and interpretations of the data, which are discussed in more detail below.

Music Perception

Quantifying music perception is an interesting test case for this kind of data gathering and analytical paradigm. Most music or auditory perception studies have the inherent confound that small changes can affect listeners' perception, especially when the study involves timing, tuning, or sound localization. However, the experimental controls may be loosened slightly when investigating holistic music listening, because no single musical element is more important than the whole.

Quantitative and qualitative elements of music are theoretically distinct but practically inseparable (Bruner II, 1990). Listeners respond affectively to technical aspects of music using schemata informed by their individual musical experiences and personality traits (Kopacz, 2005), and composers use various musical and compositional techniques to convey the emotions they want to express (Battcock & Schutz, 2019; Bruner II, 1990).

However, quantifying the perceptual interactions between more than one or two musical qualities is a challenge. One reason is that models like ANOVA and its variations are limited by how many variables a researcher can include while remaining coherent. Another is that asking participants to respond to multiple aspects of a stimulus taxes participants' perceptual capacity and is thus difficult to measure (W. F. Thompson, 1994).

One specific area that has attempted to capture a greater dimensionality is music emotion research. This is a well trod domain — see, for example Juslin and Sloboda (2010) — and the application of multivariate analyses to these questions is similarly well established. Early studies, including Gray and Wheeler (1967), Wedin (1969), and Wedin (1972) used MDS to capture the affective space of various musical stimuli. MDS continues to be used commonly in more modern studies (Bigand et al., 2005; Madsen, 1997; Rodà et al., 2014), with a narrow focus on valence and arousal, which were first proposed to be the most salient dimensions of perception by Osgood and Suci (1955).

A few studies have specifically investigated dimensions beyond those first two (for example Rodà et al., 2014), and there are conflicting hypotheses about whether the valence-arousal plane a different model of emotion perception represents the fundamental space around music emotion perception (Cowen et al., 2020; Juslin & Västfjäll, 2008). However, an important distinction between the present study and work in music emotion perception is that the adjectives we chose were informed by music composition and performance, rather than emotion (Wallmark, 2019).

With regard to trained and untrained musicians, there are many studies that evaluate their differences, but the verdict as to whether trained musicians are better music listeners is still out, partially due to the fact that there is little standardization in how much training is required for a participant to be "highly trained" (Bigand & Poulin-Charronnat, 2006). There are, however, reported benefits with regard to sensitivity to the emotional content in music (Ladinig & Glenn Schellenberg, 2012) and familiarity with tonal systems

(Bartlett & Dowling, 1980; Dowling, 1978). Recent works suggest that these benefits may be limited to specific technical aspects, and depend on the extent of training (Raman & Dowling 2017). We included highly trained musicians because they are sensitive to these technical aspects of music and will be able to accurately quantify the stimuli. Additionally, some of the response options to questions on the survey for Experiment 1 would only be familiar to participants with significant music training.

Intercultural music perception.   There are a few common goals in intercultural studies of music perception. Some quantify the shared emotional experience between musical cultures (Balkwill et al., 2004; Balkwill & Thompson, 1999; Cowen et al., 2020; Darrow et al., 1987; Fritz et al., 2009; Gregory & Varney, 1996), and some ask participants to identify technical aspects of music from other cultures (Raman & Dowling, 2016, 2017). There are fewer studies that include semantics in their evaluation of music perception (Zacharakis et al., 2014, 2015), which makes this a prime area for research.

The research program presented in Zacharakis et al. (2014) and Zacharakis et al. (2015) deals specifically with timbre perception, and their use of adjectives is similar to the way they adjectives are used in the present study. In Zacharakis et al. (2014, 2015), Greek and English participants described timbre with adjectives from their native languages. These studies found that while there are some differences, overall, participants' descriptions of timbre do not differ much between languages (Zacharakis et al., 2014, 2015).

Present questions & methods of analysis

The primary question addressed in this study is: Can we quantify a cognitive space around music listening defined by both stimulus and cognitive dimensions of music. Secondary questions include whether French and American participants describe music differently, and whether those differences may arise from cultural differences or are purely semantic. To answer these questions, we employed a set of multivariate analyses that each offered a different perspective on the results of each experiment. We felt it may be useful to

provide a quick overview of the data collection and analytical techniques for readers who may be unfamiliar.

Check-all-that-apply (CATA).   The CATA technique — a method widely used in sensory evaluation — measures how participants describe a set of stimuli (Coombs et al., 1956; Katz & Braly, 1933; Meyners & Castura, 2014; Valentin et al., 2012). In a CATA task, stimuli are presented one at a time, and for each stimulus, participants are shown a list of descriptors and are asked to select the descriptors that describe the presented stimulus (Meyners & Castura, 2014). CATA easily assesses questions with multiple 'correct' responses (Coombs et al., 1956), and places little cognitive demand on participants because they do not have to generate responses (Ares et al., 2010).

CATA data are analyzed by 1) computing a pseudo contingency table that records the number of times descriptors were associated with stimuli and 2) analyzing this data table with Corresponence Analysis (CA) in order to visualize the patters of association between a) stimuli, b) descriptors, and c) stimuli and descriptors.

Correspondence Analysis.   The primary analysis used on the data collected in the surveys is Correspondence Analysis (CA) (Benzécri, 1973; Escofier-Cordier, 1965; Greenacre, 1984). CA analyzes a contingency table, or any data structured similarly, and calculates the relationships between rows (observations) and columns (variables); in our case, musical excerpts and descriptors. The nature of the CA allows for observations and variables to be visualized in the same space.

Hierarchical Cluster Analysis.   Hierarchical Cluster Analysis (HCA) (Pielou, 1984) is a way of identifying groups, or clusters, of observations from the rows of a distance matrix, that displayed using trees to visualize that distance. This technique was used in the present study to determine whether there were any clusters of excerpts or adjectives that arose during participant ratings. These clusters were used as design or grouping variables and to select colors for visualizations.

Multidimensional Scaling.   Multidimensional Scaling (MDS) (Borg & Groenen, 2005) — a technique commonly used in music perception studies (Bigand et al., 2005; Madsen, 1997; Rodà et al., 2014; Wedin, 1969, 1972) analyzes a distance matrix computed between observations and visualizes them, positioning these observations on a map such that the distance between observations on the map best approximates their distance in the data table.

Multiple Factor Analysis.   Multiple Factor Analysis (MFA) (Abdi et al., 2013) is an extension of PCA that analyzes and visualizes multiple tables or sets of variables that each describe the same observations. MFA visualizations, called are focused on the relationships between observations, and, for each observation, the relationships between the tables that contributed to that observation. These visualizations represent *partial factor score plots*, such that "partial" refers to each table's contributions to the overall factor scores. Practically speaking, the most basic difference between MFA and PLSC is that PLSC extracts commonalities between two data tables, whereas MFA extracts similarities and differences between two or more data tables.

Partial Least Squares Correlation.   Partial Least Squares Correlation (PLSC) (Abdi & Williams, 2013; Tucker, 1958) analyzes two data tables that describe a single set of observations (rows) with different sets of variables (columns). PLSC computes a matrix of correlations between the sets of variables which is then analyzed to find latent variables with the largest covariance, i.e., the greatest amount of information common to the two data tables. This technique is commonly used in neuroimaging studies to evaluate correlations between matrices of imaging data and of behavioral or task data (Krishnan et al., 2011).

Inference Methods.   Because the methods outlined above are not inferential methods, and do not inherently allow for hypothesis testing, we use permutation testing (Berry et al., 2011) and bootstrapping (Hesterberg, 2011). Permutation testing evaluates whether our data have a signal that is more salient than a random table and bootstrapping

evaluates the stability of the result of an experiment with confidence intervals computed from resampling the original observations (Berry et al., 2011; Hesterberg, 2011).

Experiment 1: Musical Qualities Survey

Methods

Participants. For the first experiment, we recruited highly trained musicians with a minimum of 10 years of formal music training to evaluate the stimulus dimensions or musical qualities of the stimuli, and to ascertain whether the stimuli truly reflected the composer's intent of varying on a wide range of musical dimensions (Raman & Dowling, 2017). Participants in the United States and in France were recruited by word of mouth and social media. There were a total of 84 responses to the survey, of which 57 were removed for not completing the survey, leaving a total of 27 ($N_{France} = 9$, $N_{USA} = 18$) for the analysis. All recruitment measures were approved by the UT Dallas IRB.

Stimuli. All stimuli were new, original excerpts, in various Western styles, composed using Finale composition software (Finale v25, MakeMusic, Inc.) by the first author specifically for this study (scores and audio files available upon request). Each stimulus was a wav file generated using the Finale human playback engine, approximately 30 s in length (range: 27 - 40 s, $M = 32.4$ s). The stimuli were all string quartets, in order to control for effects of timbre and otherwise vary on a number of musical qualities, specifically Harmony, Tempo, Meter, Density, and Genre. The stimuli are numbered by the order in which they were composed and thus do not reflect any specific grouping.

Survey. American and French participants received links to surveys presented in English and French, respectively, both presented via Qualtrics. Participants were instructed to listen to the listen to the excerpts presented in the survey using headphones or in a quiet listening environment, but that was not controlled, nor was it assessed as part of the survey. After standard informed consent procedures, participants listened to 15 of the 30 excerpts, presented one at time in a random order, and answered ten questions per

excerpt, one for each of the musical qualities being assessed. The musical qualities assessed and the levels associated with each are indicated in Table 1. Of these dimensions being assessed, all were multiple choice, allowing for a single response, except for meter, contour, and articulation, which were check-all-that-apply. See supplementary materials for the French translations of this table. After completion of the experimental task, participants were asked to provide demographic data, including including age, gender identity, nationality, occupation, and musical experience.

Table 1

Musical Qualities and the provided survey response options.

Harmonic Material	Tempo	Meter	Density	Genre	Articulation
Diatonic: Major	Very slow	Simple Duple	Very sparse	Baroque	Staccato
Diatonic: Minor	Slow	Simple Triple	Moderately sparse	Classical	Marcato
Blues	Moderately Slow	Simple Quadruple	More sparse than dense	Romantic	Legato
Chromatic	Moderate	Compound Duple	More dense than sparse	Impressionist	Tenuto
Whole tone	Moderately Fast	Compound Triple	Moderately Dense	Modern	Other
Modal	Fast	Compound Quadruple	Very Dense	Jazz/Blues	
Quintal/Quartal	Very Fast	Complex		Contemporary	
Ambiguous				Other	
Other					

Contour	Motion	Range	Dynamics
Ascending	Conjunct	Narrow	Soft
Descending	Disjunct	Moderate	Moderate
Arch	Combination of conjunct	Wide	Loud
Undulating	and disjunct	Very Wide	Varied: gradual crescendo
Pendulum	I do not think this excerpt has a melody	I do not think this excerpt has a melody	Varied: gradual decrescendo
Terrace	Other		Some of each, soft and loud
I do not think this excerpt has a melody			Other
Other			

Data Processing.   To process the data, survey responses were converted into a "brick" of data. With the excerpts on the rows, the qualities on the columns, and the participants on the pages. For the current experiment, one quality is one variable, and we refer to the response options as levels of that variable. On any page, at the intersection of any row and column was a one or a zero, with a one indicating that this participant had selected this level of this quality (column) to describe this excerpt (row). The responses in the French "brick" were all translated into English, and then the brick was summed across pages to obtain a pseudo-contingency table[1] in which the intersection of a row and a column was the total number participants who selected a level of a musical quality to describe an excerpt. Levels for which the column sum was equal to one were considered as outliers and removed from the data.

After removing these columns, preliminary visualizations revealed a few variables that required recoding because they were having an outsized effect on the analysis. For the "Meter," variable, there were initially seven levels: "Simple Duple," "Simple Triple," "Simple Quadruple," "Compound Duple," "Compound Triple," "Compound Quadruple," and "Complex," but some participants misunderstood this question and selected multiple options for each level of "Duple," "Triple" or "Quadruple." These responses were recoded, removing "Simple," "Compound" and "Complex," (there were no excerpts with complex meter), and collapsed, so that each excerpt had only one meter response per participant, "Duple," "Triple," or "Quadruple."

There were multiple qualities for which a possible response was "I do not think this excerpt has a melody," and this pattern created a problem in which multiple columns represented the same response, which had a similar effect to the one caused by the "Meter"

───────

[1] Whereas, in a contingency table, there is a single one (1) in each row, indicating that each observation (row) is associated with a single variable (column), in a pseudo-contingency table, there are as many ones as variables the participant decided were associated with that observation.

variable before it was recoded. To avoid this problem, these responses were also recoded. A new variable, "Melody," was created, with two levels/columns, yes and no, and if participants responded "I do not think this excerpt has a melody" to any of the Contour, Motion, or Range variables, a one was counted in the "No" column for that participant and that excerpt. The other levels for each of these three variables were then recoded so that each other column for that variable in that row had the value of one divided by the number of options for that variable — a procedure called barycentric recoding. If the participant responded with "I do not think this excerpt has a melody" for some but not all three of those variables, a one was still counted in the "no melody" column, but only the variables for which "I do not think…" was selected were recoded using barycentric recoding. For all excerpts and participants for which "I do not think…" was never selected, a one was added to the "Yes" column for the melody variable. Once the data were recoded, the brick was once again summed across pages to obtain the data table that would be used for subsequent analyses.

Analysis. After data processing, the data were in two usable forms, a brick and a contingency table. To analyze the similarity structure between participants, we computed a co-occurrence matrix from the brick with participants on the rows and columns, such that the intersection of a row and column represented the number of similar choices between participants. This was then analyzed using MDS.

To analyze the excerpts and musical qualities and obtain the music quality space, we performed a CA on the contingency table. To identify potential clusters among the excerpts, we ran an HCA on the row factor scores obtained from the CA.

Results

Participants. The MDS performed on the co-occurrence matrix of participants was intended to identify potential clusters of participants. Visual examination of the results did not reveal any clusters — a pattern suggesting that the participants constituted

a homogeneous group. To further assess this conclusion, we also computed average factor scores by nationality and gender identity and bootstrap-derived confidence intervals around these averages and did not find any significant differences. See supplementary materials for plots illustrating this.
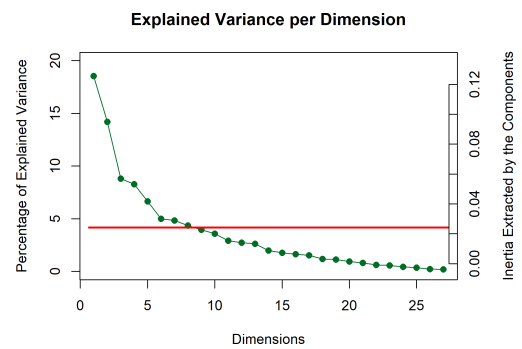
Excerpts.

The results of the CA performed on the contingency table revealed two significant dimensions, accounting for 32.74% of the total variance. Figure 1 displays the scree plot, which shows for this analysis the percentage of variance explained by each dimension.

Figure 1. Scree plot for the CA of the Musical Qualities Survey

Note. The x-axis represents dimensions and the y-axis represens the amount of variance extracted by each dimension.

Preliminary plots of the factor scores obtained from the CA revealed that Excerpts 6 and 14 distorted the factor space, with these two excerpts dominating the second and third dimensions, respectively. To help interpret the factor space, these two excerpts were removed from the data and the CA was recomputed. Excerpts 6 and 14 were then added back in as *supplementary observations*, a technique which visualizes the information these elements share with the other elements in the sample without distorting the factor space. The proximity of these supplementary observations to the origin helps identify how much information is shared with the rest of the sample. The closer the supplementary observations are to the origin, the less information they share with the rest of the sample, and if they appear directly on the origin, that suggests that they share no information.

The HCA revealed four clusters (see supplementary materials for the dendrogram). Figure 2a displays these clusters represented by colors, with Excerpts 6 and 14 as supplementary observations colored separately. Figure 2b displays the qualities, with the

levels of a given quality colored the same. For clarity, only the levels of qualities that contributed significantly (see below, and Figure 3) are displayed.

The proximity between any two points in Figure 2 can only be interpreted directly as similarity when the points are on the same map. Proximity between points on separate maps must be interpreted relative to the center of mass. Because CA captures "average" information, greater proximity between an excerpt and the level of a quality can be interpreted as the two being associated with one another *more than the average* (Abdi & Williams, 2010).

Figure 2. Factor Plots for Rows and Columns of the Qualities Survey
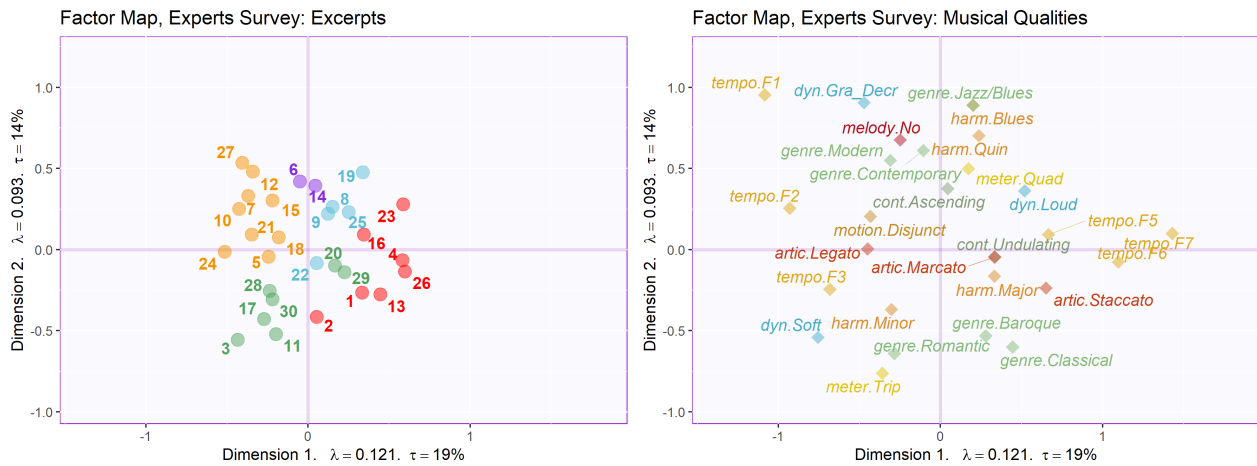
Note. Factor scores for the CA of the QS. Axes are labeled with the dimension, the eigenvalue, and the explained variance. Row and column factor scores displayed separately. Only levels of variables that contribute significantly to the first two dimensions are plotted.

To evaluate the relative importance of the excerpts and musical qualities in defining each dimension, we computed the *contributions* of the excerpts and qualities to the factorial dimensions. Contributions are similar to a squared coefficient of correlation. They vary between zero and one with zero indicating no importance and one indicating maximum importance (Abdi & Williams, 2010). Contributions—being squared—are

positive, but to facilitate interpretation contributions are signed to express the sign of the factor scores. Contributions whose magnitude is larger than the average contribution are considered important for their factorial dimensions. A plot of the contributions for all variables is in the supplementary materials.

Figure 3 shows the contributions of excerpts and qualities important for first two dimensions. For Dimension 1, Excerpts 4, 13, 23, and 26, contribute to the positive side of the dimension, whereas Excerpts 3, 7, 10, 24, and 27 contribute to the negative side of the dimension. Low tempi (tempo.F2 and tempo.F1) contribute to the negative side, along with legato (smooth) articulation and soft dynamics. Finally, high tempi contribute to the positive side, along with marcato and staccato (accented and separate, respectively) articulations and loud dynamics. Single levels from other variables also contribute to the first dimension: major harmony, classical genre, and undulating contour in the positive direction and disjunct motion and triple meter in the negative direction.

For Dimension 2, Excerpts 7, 12, 15, 27, and 29 contribute to the positive side of the dimension, whereas Excerpts 2, 3, 11, and 17 contribute to the negative side. Newer genres (Contemporary, Jazz/Blues, and Modern) contribute to the positive side of Dimension 2, along with Blues and Quintal harmony, Quadruple meter, Gradual decrescendo and loud dynamics, ascending contour, "no melody," and the slowest level of the tempo variable. Older genres (Baroque, Classical, and Romantic) contribute to the negative side of this dimension, along with soft dynamics, minor harmony, and triple meter. A full enumeration of contributions and bootstrap ratios is available at the URL in the author note.

Discussion

Figure 2 reveals a few musical connections: for example between tempo and articulation (on Dimension 1) is apparent, and between genre and harmony (on Dimension 2). Staccato articulations, associated on this factor plot with high tempi, are played light

Figure 3. Important contributions for the CA of the Musical Qualities
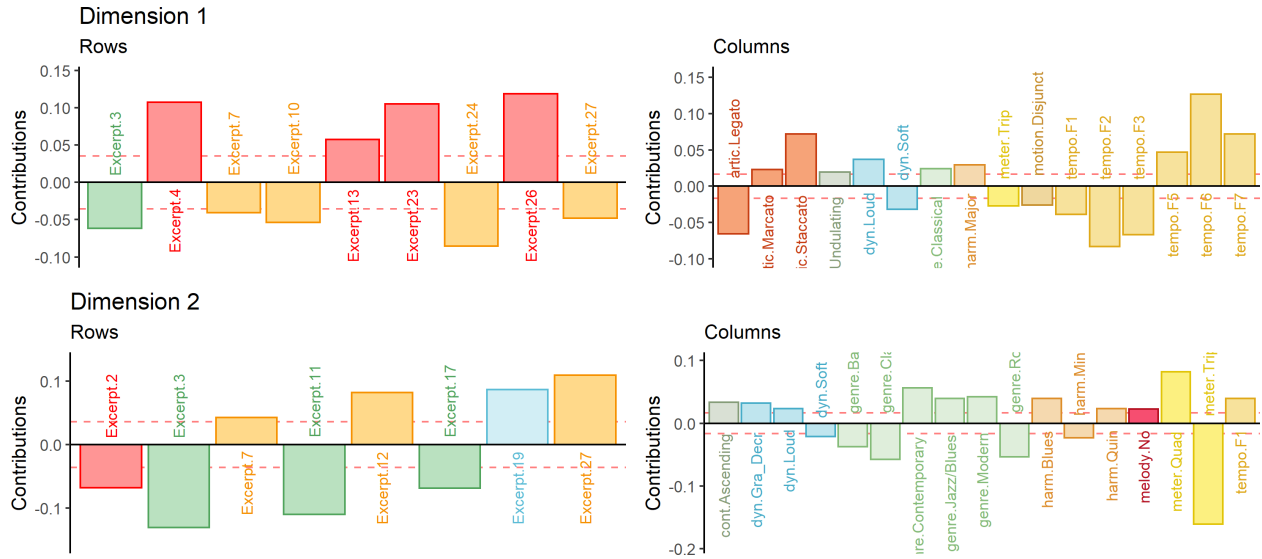
Note. The contributions plotted here are the excerpts and levels of variables with contributions important to Dimensions 1 and 2 plotted with magnitude and direction. The y-axis represents the value of the contributions. The excerpts are colored according to the clusters extracted by the HCA and the levels of each variable are all colored the same.

and separate, and legato articulations, associated with slow tempi, are played smooth and connected. In terms of performance practice, slow or long notes are played in a legato style to create a sense of continuity, and fast moving notes or phrases do not require the same technique. The coordinate mapping of jazz/blues harmony and genre, which are on top of one another is the most extreme example of a genre being associated with certain harmonic material, but other connections are also revealed. The second dimension separates the older styles, such as Baroque, Classical, and Romantic, from newer styles, Contemporary, Jazz/Blues, and Modern. Dimension 2 associates the simpler harmonies of major and minor and the more complex harmonies of quintal and blues with older and newer styles, respectively, following historical practice. The Classical era had fairly structured rules for both harmony and voice leading, but the Romantic era relaxed those rules and introduced

more complex harmonies (Cohn et al., 2001). The gradual devolution of those rules and the increase in complexity of harmony continued through the modern and contemporary styles (Cohn et al., 2001; Kennedy et al., 2013).

These musical connections facilitate our interpretation of the first two dimensions. The first dimension can be interpreted as arousal—tempo, articulation, and dynamics all load from greater arousal in the positive direction to lesser arousal in the negative direction on the first dimension. Dimension 2 is less clear, and does not seem to be tied to valence. Minor and major harmony, for example, both score negatively on Dimension 2. Instead, Figure 3 shows that while two levels of the meter variable are the most important for this dimension, that genre is also important, based on the number of levels of genre that contribute significantly to Dimension 2. Considering these and the harmony variables, it may be that the second dimension represents complexity.

Finally, we note that because of the nature of this survey, these results tell us more about the excerpts themselves rather than the behavior of the participants. The participants rated the stimuli similarly, validating the variety among the excerpts, indicating that they are different enough to create a large and varied factor space.

Experiment 2: Musical Adjectives Survey

Methods

Participants.   Participants with self-reported normal hearing were recruited for Experiment 2 without regard to level of music training. Participants in the United States were recruited through the UT Dallas Psych Research Sign-up (SONA) System and by word of mouth and social media. French participants were recruited by word of mouth, email, and social media. Only participants who signed up via the SONA System were compensated with research participation credit, other participants were not compensated in any way. Of a total 520 survey responses, 160 were removed for being incomplete,

leaving a total of 354. Additionally, participants from the US who indicated solely a nationality other than American were excluded from analysis. "Ghanian," for example, was not included, but responses such as "Indian-American" were. This left a total of 178 ($N_F =$ 112, $N_A = 166$) survey responses for analysis. All recruitment measures were approved by the UT Dallas IRB.

Stimuli.   The stimuli used for Experiment 2 are the same as those used for Experiment 1.

Survey.   The procedure for participants in Experiments 1 and 2 was similar. They received a link to a Qualtrics survey presented in either English or French, depending on their location, and instructions regarding listening environment were the same. After standard informed consent procedures, the participants listened to 15 of the 30 excerpts, presented one at a time, in a random order, and performed a CATA task. The task was to evaluate the stimuli using thirty-three adjectives such as 'dark,' 'warm,' and 'colorful' (French: 'sombre,' 'chaleureux,' and 'colore'). The adjectives for the AS were selected using Wallmark (2019) as a guide and in consult with a French professional musician. Some adjectives were initially selected in French and some in English. In all cases, adjectives were selected for which there was a clear French (vis-à-vis English) translation. The adjectives are listed in English and in French in the supplementary materials. Following the experimental task, the participants were asked to provide demographic data, including age, gender identity, nationality, occupation, and musical experience.

Data Processing & Analysis.   Data for the survey for Experiment 2 (Adjectives survey/AS) were processed similarly to those for Experiment 1. Due to a technical error, French participants were not presented with Excerpt 17, so the data from that excerpt was removed from the dataset for the American participants. To process the data, first, all French survey responses were translated into English. Both sets of responses were then converted into "bricks" of data, with the excerpts on the rows, the adjectives on the columns, and participants on the pages. On any page, at the intersection of any row and

column was a one or a zero, with a one indicating that this participant had selected this adjective (column) to describe this stimulus (row). The bricks were then concatenated and summed across pages to obtain a pseudo-contingency table in which the intersection of a row and a column was the total number participants who selected an adjective to describe an excerpt.

After data processing, the data were in two usable forms, a brick and a contingency table. To analyze the similarity structure between participants, we computed a co-occurrence matrix from the brick with participants on the rows and columns, such that the intersection of a row and column represented the number of similar choices between participants. This was then analyzed using MDS.

To analyze the excerpts and adjectives and obtain the music quality space, we performed a CA on the contingency table. To identify potential clusters of excerpts or adjectives, two separate HCAs were computed, one each on the row factor scores (excerpts) and the column factor scores (adjectives) obtained from the CA.

We used MFA to explore the differences between the ratings of French and American participants. To prepare the data, we summed separate pseudo-contingency tables for the French and American participants, with the excerpts on the rows and the adjectives on the columns. Because MFA does not evaluate rows and columns in the same analysis the way the CA does, we had to transpose each of these tables to obtain tables with the adjectives on the rows and the excerpts on the columns. The two pairs of tables were concatenated and we then ran separate MFAs on these.

Results

Participants. Figure 4 displays the factor scores obtained from the MDS performed on the co-occurrence matrix for the AS, along with group means and bootstrap-derived confidence intervals. The separation between the confidence intervals indicates significant

differences between French and American participants, $p. < .01$. Additional analyses using gender identity and level of music training as factors were not found to be significant.

Figure 4. R_V Analysis of Participants in the Adjectives Survey
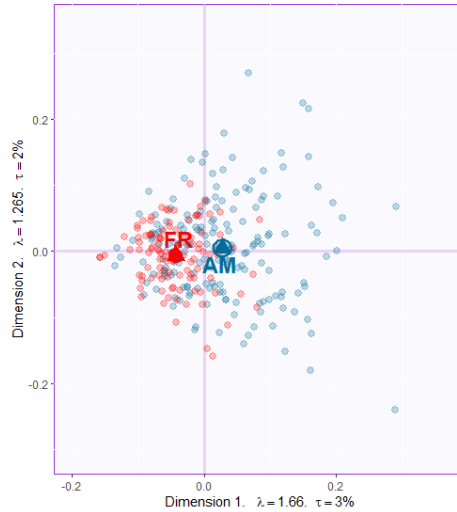
Note. Axis labels are the same as Figure 2. Factor scores for French and Americans are colored by group, in blue and red, respectively. Group means are indicated with triangles and labled with AM and FR. The ellipse around the group mean indicates the confidence interval, after bootstrapping 1000 iterations. The clear separation between the confidence intervals suggests there is a significant difference between the groups, $p > .001$.

Excerpts. The CA performed on the AS pseudo-contingency table revealed two significant dimensions, accounting for 72.55% of the variance total variance. Figure 1 displays the scree plot, which shows for this analysis the percentage of variance explained by each dimension. Although excerpts 6 and 14 were removed from Experiment 1 data for being outliers, they were not outliers in this analysis, and were therefore included in all of the analyses for Experiment 2.

Figure 5. Scree plot for the CA of the Adjectives Survey
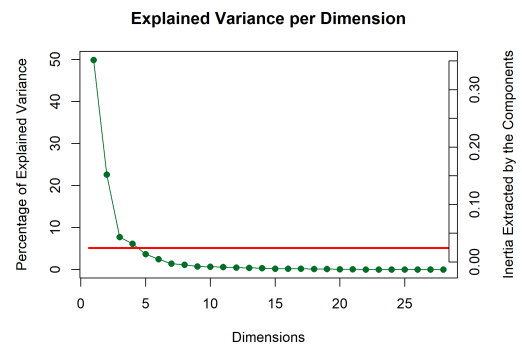
Note. This plot shows the percentage of explained variance per dimension for this analysis. The horizontal line is called the Kaiser line and indicates the average vari-

Figure 6. Factor Plots for the CA of the Adjectives Surveys
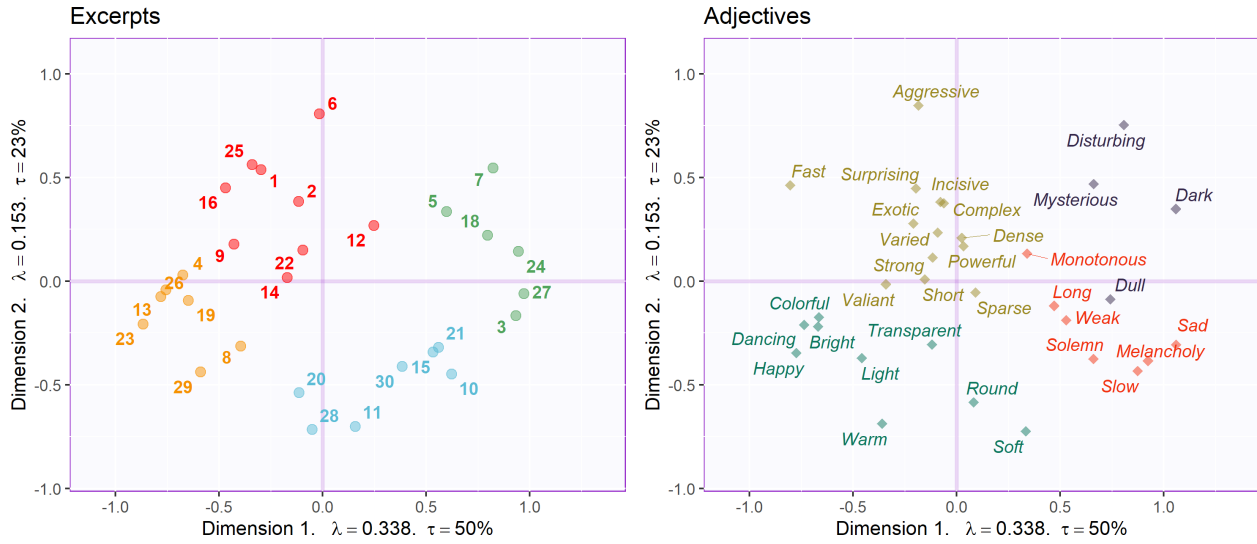
Note. This plot shows the row and column factor scores displayed separately. Excerpts and Adjectives are colored according to the clusters exctracted by their respecitve HCAs

The HCAs computed separately on the row and column factor scores revealed four clusters for each, and the excerpts and adjectives are colored according to those clusters in Figure 6. This figure displays the row (excerpt) and column (adjective) factor scores obtained from the CA computed on the AS contingency table. Interpretation of the maps individually and jointly is the same as for Experiment 1.

Figure 3 shows the contributions of excerpts and qualities important for first two dimensions. For Dimension 1, Excerpts 4, 13, 23, and 26, contribute to the positive side of the dimension, whereas Excerpts 3, 7, 10, 24, and 27 contribute to the negative side of the dimension. Low tempi (tempo.F2 and tempo.F1) contribute to the negative side, along with legato (smooth) articulation and soft dynamics. Finally, high tempi contribute to the positive side, along with marcato and staccato (accented and separate, respectively)

articulations and loud dynamics. Single levels from other variables also contribute to the first dimension: major harmony, classical genre, and undulating contour in the positive direction and disjunct motion and triple meter in the negative direction.

For Dimension 2, Excerpts 7, 12, 15, 27, and 29 contribute to the positive side of the dimension, whereas Excerpts 2, 3, 11, and 17 contribute to the negative side. Newer genres (Contemporary, Jazz/Blues, and Modern) contribute to the positive side of Dimension 2, along with Blues and Quintal harmony, Quadruple meter, Gradual decrescendo and loud dynamics, ascending contour, "no melody," and the slowest level of the tempo variable. Older genres (Baroque, Classical, and Romantic) contribute to the negative side of this dimension, along with soft dynamics, minor harmony, and triple meter. A full enumeration of contributions and bootstrap ratios is available at the URL in the author note.

Figure 7 displays the contributions of excerpts and adjectives important for the first two dimensions. For Dimension 1, Excerpts 3, 7, 10, 18, 24, and 27 contribute to the positive side of the dimension, whereas Excerpts 4, 13, 19, 23, and 26 contribute to the negative side. Adjectives "Dark," "Disturbing," "Melancholy," "Mysterious," "Sad," "Slow," and "Solemn" contribute in the positive direction, whereas "Bright," "Colorful," "Dancing," "Fast," and "Happy" contribute in the positive direction.

For Dimension 2, Excerpts 1, 6, 7, 16, and 25 contribute to the positive side of the dimension, whereas Excerpts 10, 11, 20, 28, and 29 contribute to the negative side. Adjectives "Aggressive," "Complex," "Disturbing," "Fast," "Mysterious," and "Surprising" contribute to the positive side of Dimension 2, whereas "Happy," "Melancholy,""Round," "Slow," "Soft," "Solemn," and "Warm" contribute to the negative side.

Figure 8 displays partial factor score plots (see "Multiple Factor Analysis" above) of the results of the MFA evaluating the differences in rating behavior between French and American participants. Although the specific magnitudes of the dimensions extracted by

Figure 7. Contribution Barplots for Rows and Columns of the Adjectives Surveys
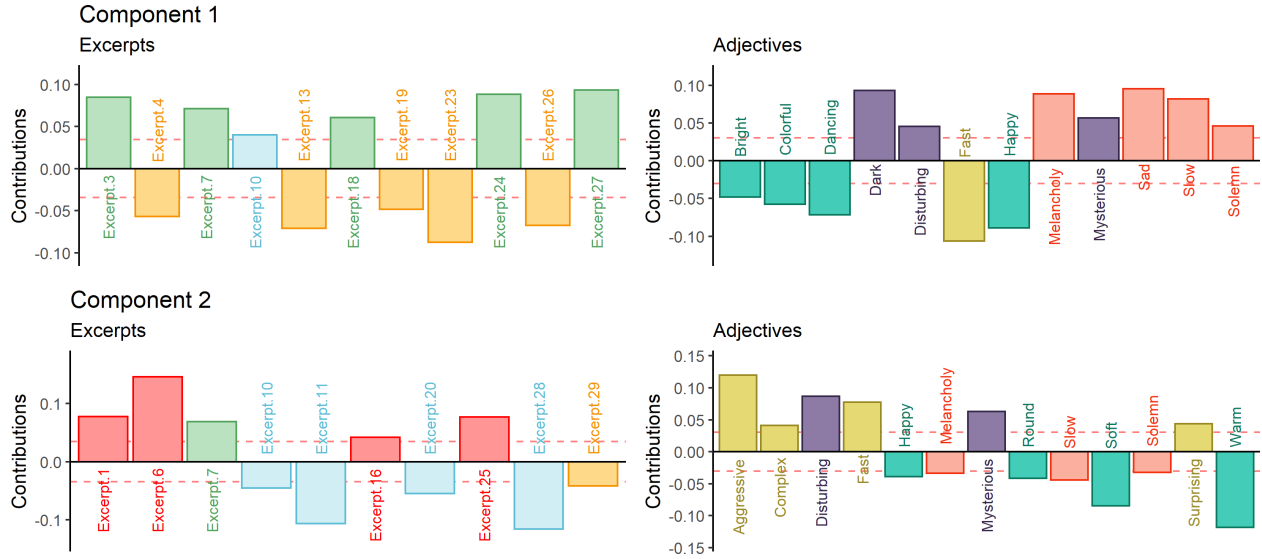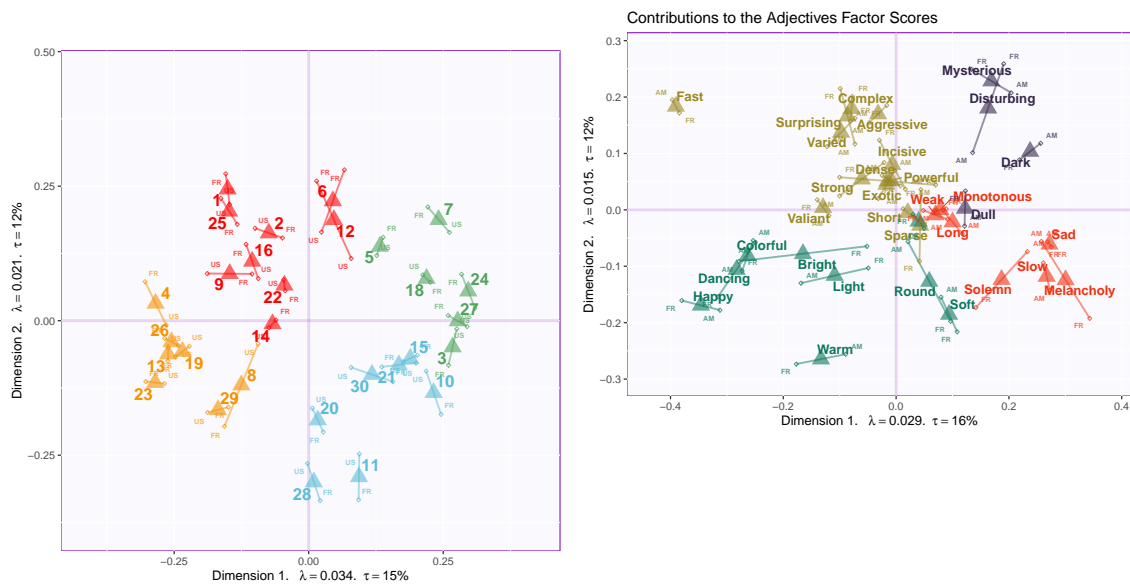

Note. The excerpts and adjectives with contributions important to Dimensions 1 and 2 are represented by their magnitude and direction. Both excerpts and variables are colored according to the clusters extracted by their respective HCAs.

the MFAs are different, as evidenced by the percentage of variance extracted (τ) on the axes of the plots, the general space revealed by these analyses is similar to that revealed by the CA. Thus we can interpret the space similarly, relative to the valence-arousal plane. However, in this case, we cannot compare elements between maps.

The triangles represent the compromise between the mental spaces of the French and American participants for each item, and the lines extending from the triangles indicate the partial factor scores for the items from the perspective of each group (Abdi et al., 2013). Excerpts and adjectives that were rated similarly have shorter lines extending from them, but excerpts that were rated differently have much longer lines. Examples of excerpts that were rated differently are numbers 6, 8, and 12. Adjectives that were used differently include "Disturbing," "Round," "Solemn," and "Bright."

Figure 8. Partial Factor Scores Plots from the MFA on the Adjectives Survey

(a) For the Excerpts.   (b) For the adjectives.

Note. In each plot, the triangles represent the combined factor scores and the small circles represent the partial factor scores contributed by each of the groups. Note the different values between plots in the dimension labels, indicating separate analyses.

Discussion

There are clues to the nature of the group differences between French and American participants in Figure 4, the factor scores plot for the MDS. The greater variance among American participants, represented by the greater spread on the plot, highlights a much less homogeneous response to the excerpts relative to the French participants. This may be due to a much less homogeneous set captured by the American population sample. Among the American sample, there were a number of responses in which participants identified with another nationality in addition to American. However, all of the French participants responded with "French," except for one participant, who responded with "French - Belgian."

Although this experiment was designed to evaluate the cognitive response to music, and not the emotional response, there is significant overlap in the results observed here and the results of the work investigating music and emotion, specifically in the appearance of the valence-arousal plane. Studies on music and emotion have used this construct as a way of measuring emotion, but the original proposal was that this plane is simply a measure of "meaning" (Osgood & Suci, 1955). The adjectives selected for use in Experiment 2 reflect this original proposal, as they were selected specifically to represent a cognitive space, not an emotional one.

Interpreting the differences in how the adjectives and the excerpts are distributed in the space is aided by considering the valence-arousal plane. Excerpts 6 and 12 tell a similar story, that the French participants tended to assign adjectives with higher arousal to these excerpts than the American participants. Excerpt 8 is similar, but with low arousal. More interesting, perhaps, is the difference between participants in rating Excerpt 4. Although the magnitude of the difference isn't as extreme, the American partial factor score lies just below the x-axis, in the positive valence/low arousal quadrant, and the French partial factor score lies above the x-axis, in the positive valence/low arousal quadrant.

Adjectives that highlight the differences between French and American participants include "Bright," "Light," "Round," "Solemn," "Melancholy," "Mysterious," and "Disturbing." "Bright" (Brillant) may be the most extreme example, as the French partial factor score is close to the origin and the American partial factor score is much further away, suggesting that this word carries much more positive valence in English than in French. Additionally, The location of the American partial factor score for "Bright" suggests that Americans commonly grouped it with "Colorful" and "Dancing," a contrast to how the French participants used it. In English, "melancholy" (melancolique) and "sad" (triste) were used the same, but in French they were used very differently. "Disturbing" (inquiétant) is much further from the origin, suggesting that it is more extreme for the French participants than the Americans. The location of "Solemn" (Solennel) suggests that it carries more valence in English, but more arousal in French. All of these differences highlight possible differences in semantic associations, and frequency of use between languages.

Experiment 3: Combined Surveys

Methods

Experiment 3 was a Partial Least Squares Correlation performed using the pseudo-contingency tables created for Experiments 1 and 2. Tables used in PLSC require the same set of observations, so the data for Excerpts 6 and 14 was removed from the table for Experiment 2 and the data for Excerpt 17 was removed from the table for Experiment 1. The goal of this experiment was to identify the strongest shared signal between the two tables. PLSC shows what is common between two different sets of data—how often an excerpt was associated with *both* a musical quality and an adjective. The visualizations below show which adjectives are associated with which musical dimensions, and how the excerpts are grouped in a joint factor space. The dimensions extracted from the each table are called *latent variables* (LVs) and are shown with the latent variables of one table
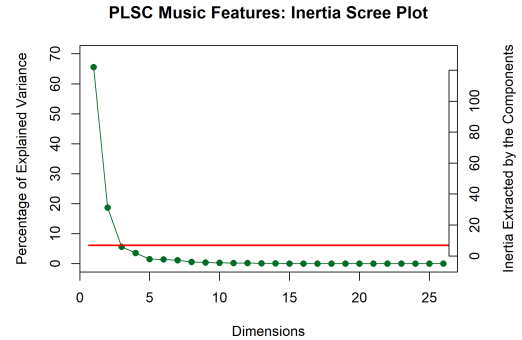
plotted against the latent variables of the other.

Results & Discussion

The PLSC performed using
the pseudo-contingency tables from Experiments
1 and 2 revealed two significant dimensions that
extracted the majority of the variance (84.25%).

Figure 11 displays the factor score
plots for this analysis with the first LVs from each
table on the left and the second LVs extracted from
the each table on the right. The excerpts are colored
according to the clusters identified by the HCA
for Experiment 2, with tolerance intervals plotted
containing the excerpts in each group to highlight
their distribution. The two sets of latent variables

Figure 10. Scree plot for the PLSC

Variance extracted per dimension,
with two significant dimensions
above the Kaiser line, in red.

extracted by the analysis separate the groups of excerpts approximately by the clusters

defined in the HCA for the adjectives survey. This suggests that the excerpts that are more

distant from the origin in the first LV plot are defined more by valence than arousal, and

those clustered further from the origin on the second LV plot are defined more by arousal

than valence. The tolerance intervals for the groups highlight how the groups furthest from

the barycenter in LV plot 1 are closer to the origin in LV plot 2, and the quadrants they

occupy have switched. The other two groups remain in the same quadrants between plots,

but move towards the origin. Specific excerpts that highlight these differences include

Excerpts 26, 24, and 27. Excerpt 26 is one of the most extreme examples of positive

valence, but is much closer to the origin in the second LV plot, similarly with Excerpts 24

and 27, but with negative valence. This is contrasted with Excerpt 7, which is one of the

most negatively valenced stimuli, appearing far from the origin in the first LV plot, but also
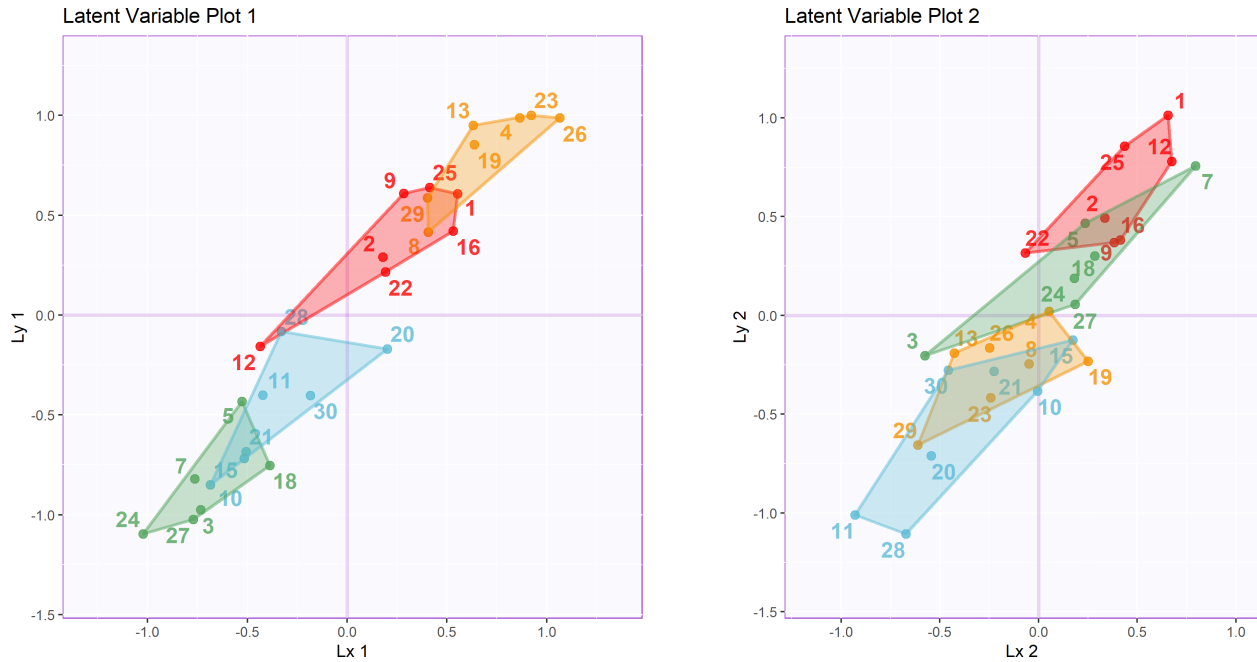
scores very high on arousal, and appears far from the origin on the second LV plot as well.

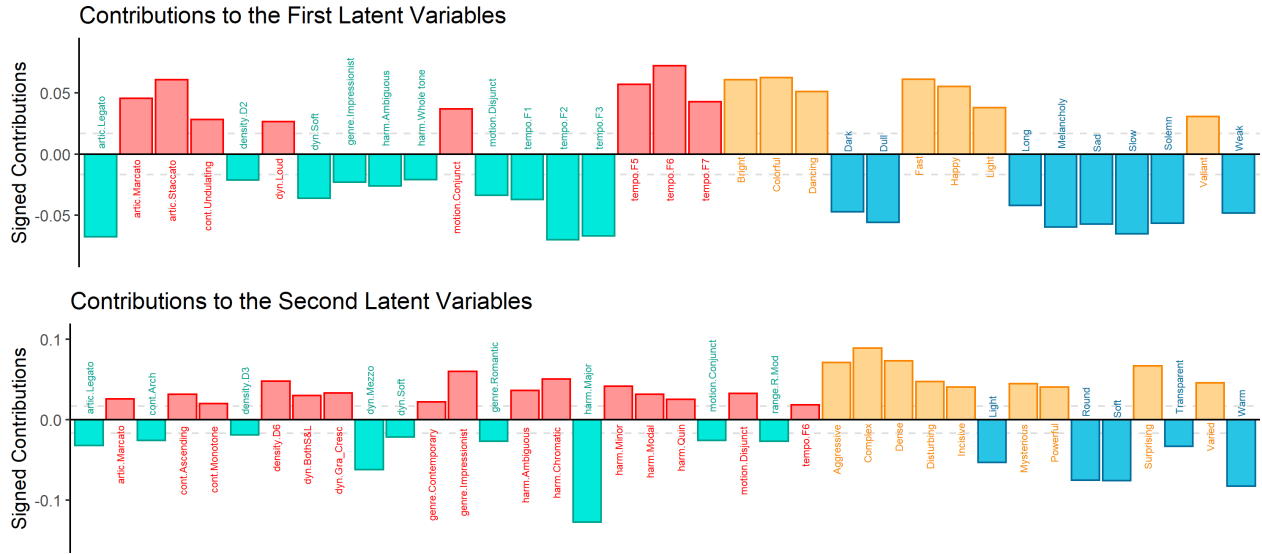*Figure 11.* Factor Plots for Latent Variables Exctracted by the PLSC

Note. Left: first latent variables extracted from the pseudo-contingency tables of the QS and the AS. Right: Second latent variables extracted from the pseudo-contingency tables of the QS and the AS. For both plots, x-axis is the LV from the QS and the y-axis is the LV from the AS.

Figure 12 displays the variables from each data table that are important for the first and second latent variables. For this plot, the important levels of variables from the QS are displayed in red (positive) and green (negative) and the important adjectives from the AS are in yellow (positive) and blue (negative). The first LVs associate the levels of variables identified as contributing to an arousal dimension in Experiment 1 with the adjectives identified as contributing to a valence dimension in Experiment 2. Likewise the second LVs associate the adjectives identified as contributing to an arousal dimension in Experiment 2 with many of the levels of variables identified as contributing to the genre or complexity dimension in Experiment 1. This suggests that many of the musical qualities that are associated with arousal, such as tempo, dynamics, and articulation, contribute to musical valence. Figure 6 provides a useful key for interpretation. The excerpts appear to be

separated approximately by quadrant, and less specifically by axis. This suggests that each group of excerpts reflects elements of both valence and arousal.

Figure 12. Factor Plots for Latent Variables Exctracted by the PLSC

Note. Signed contributions important for the first and second latent variables extracted by the PLSC. The line representing the average contribution is in gray. The y-axis represents the magnitude of the contribution. Contributions are colored according to the table they are from. Red and green are from the QS and yellow and blue are from the AS.

General Discussion

We collected data in the form of responses to musical stimuli and used multivariate analyses to explore the musical and cognitive spaces created by participants from France and the United States. The results revealed a clear valence-arousal plane that was common to participants from both countries, but significant differences in the behavior of French and American participants when responding to the stimuli using adjectives, and a space defined by arousal and complexity when evaluating the stimuli using musical qualities. The combined musical and cognitive spaces identified which musical qualities were associated with which descriptors in terms of valence and arousal.

One effect observed in Experiment 1 which was not observed in Experiment 2, in which two individual excerpts—numbers 6 and 14—dominate the factor space, requires more explanation. In a CA, information that is common falls towards the origin—the center of mass, or *barycenter*—of the factor plot, while information that is further from the average, or more rare, ends up further away (Abdi & Williams, 2010). Therefore, if any individual item on a survey is rated significantly different from the rest of the items, that item can dominate the factor space.

In this case we have two examples of stimuli that were different enough from the others in the set that they had this effect. Excerpt 6, which was a Steve-Reich-esque, minimalist, ostinato based excerpt, and Excerpt 14, a jazzy excerpt, were both the only representative of their respective styles, and were therefore different enough from the average that they distorted the factor space. However, this effect does not appear in Experiment 2 because the effect also depends on the specific questions asked. The QS was designed to separate the stimuli along specific musical dimensions, but the AS was designed to evaluate the excerpts more generally on holistic qualities. Thus, we can compare musical stimuli from different styles directly using these holistic qualities, but the relationships between styles become more difficult to quantify as we include specific musical qualities in the evaluations.

One takeaway from these results is that a deep understanding of the stimuli may help to predict the approximate dimensionality of the solution factor space, and when designing surveys or stimuli, multiple items per group, or presumed dimension, are needed. The outliers distorting the factor space that were observed in the present study are analogous to the single noisy dimension described in the introduction. The differences in the behavior of participants between experiments in rating Excerpt 6 shows how survey design affects the overall results. It is possible that Excerpt 6 was the source of noise in Experiment 2—in addition to the effect observed in Experiment 1—due to the possibility that

untrained participants are likely to be less familiar with minimalism than the trained participants in Experiment 1, but the results are robust to that noise because the participants were not asked to rate the excerpts on any explicit dimensions or qualities.

The significant results for the participants in Experiment 2 that were not observed in Experiment 1 suggest that this experimental paradigm works as intended. Significant differences between the experts' ratings of the stimuli would have indicated that the experts were inconsistent and thus not reliable raters of the music. The significant difference in ratings between the French and American participants on the AS indicates that language connotations and word use frequency in a language do have a significant impact on how participants rate the stimuli. MFA revealed some specific differences and highlighted further possibilities for analysis.

Limitations & future directions

Although we evaluate the scores and ratings of participants from different countries, we recognize that the issue of multiculturality is not addressed to a significant degree in this study, because France and the United States are both Western countries that share Western musical culture. To truly address this question, an experiment would need to include participants from multiple, contrasting musical cultures, with languages that are more distant than English and French. However, specific musical qualities, like harmony, may not apply or translate well to other musical cultures, because the concepts of melodic and harmonic material are not the same across all musical cultures (Cohn et al., 2001; Raman & Dowling, 2017). Therefore the specific questions included in the QS would need to be adjusted.

Although we suggest that data collected in this way have a much greater hypothetical reach, we recognize that the data collected for these experiments represent a convenience sample, and many of the participants were students. However, these limitations could be easily remedied in future studies. One question that fell beyond the scope of this study

what the source of the semantic drift between languages is. Although illustrated in Figure 8, the source of the differences between French and American participants is not entirely clear. These differences may not come from cultural aspects of music listening, but linguistic sources, including the adjectives' frequency of use in either language or the cultural associations with the words (B. Thompson et al., 2020), or even the physical characteristics of the words themselves (Reilly et al., 2012). Diving more into those questions would be a fascinating future study. Another interesting study would be to use adjectives from specific domains, to see how music maps onto different sensory spaces, for example textural words, like 'moist,' 'slimy,' 'dry,' 'puckered,' 'smooth.'

Finally, the results of this study and possible extensions, in conjunction with studies that have already evaluated music perception non-verbally, may provide insight into the way in which people without language react to music, such as nonverbal autistic people.

Conclusions

Expanding collection and analytical paradigms, and thus expanding scientific scope and perspective, has the added benefit of increasing reach. By expanding the ways in which we collect data, including developing investigative paradigms that are accessible on mobile platforms and that reduce participant demand while maintaining rigor and integrity, we are able to more readily and consistently reach a broader participant population, that might normally be excluded from everday research paradigms, specifically racially and ethnically diverse populations, poor populations, those with limited access to transportation, or who have a disability, or are immunocompromised.

The present study demonstrates such an experimental paradigm, that maintains a balance between relaxing experimental control and maintaining scientific rigor. Applying similar paradigms, and focusing on collecting large amounts of qualitative data on a large number of observations, and analyzing these data using multivariate analyses will provide new avenues for investigation.

Pairing this kind of data gathering with appropriate analysis will help maintain scientific integrity. The number of ways that exist to analyze data from a single set of experiments is considerable, and the results of each analysis illuminate different parts of the story behind the data. Not every form of analysis is appropriate in every context, but understanding how, and perhaps more importantly when, to apply an analytical technique is vital to uncovering new perspectives or insights.

References

Abdi, H., & Williams, L. J. (2010). Correspondence Analysis. In N. Salkind (Ed.), *Encyclopedia of research design.* Sage.

Abdi, H., & Williams, L. J. (2013). Partial Least Squares Methods: Partial Least Squares Correlation and Partial Least Square Regression. In B. Reisfeld & A. N. Mayeno (Eds.), *Methods in molecular biology: Computational toxicology volume II* (Vol. 930, pp. 549–579). Springer Science+Business Media, LLC. https://doi.org/10.1007/978-1-62703-059-5

Abdi, H., Williams, L. J., & Valentin, D. (2013). Multiple factor analysis: Principal component analysis for multitable and multiblock data sets. *Wiley Interdisciplinary Reviews: Computational Statistics, 5,* 149–179. https://doi.org/10.1002/wics.1246

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

Balkwill, L. L., & Thompson, W. F. (1999). A Cross-Cultural Investigation of the Perception of Emotion in Music : Psychophysical and Cultural Cues. *Music Perception: An Interdisciplinary Journal, 17*(1), 43–64. https://doi.org/10.2307/40285811

Balkwill, L. L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research, 46*(4), 337–349. https://doi.org/10.1111/j.1468-5584.2004.00265.x

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A

key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(3), 501–515. https://doi.org/10.1037/0096-1523.6.3.501

Battcock, A., & Schutz, M. (2019). Acoustically expressing affect. *Music Perception*, *37*(1), 66–91. https://doi.org/10.1525/MP.2019.37.1.66

Benzécri, J.-P. (1973). *L'analyse des données.* Dunod.

Berry, K. J., Johnston, J. E., & Mielke, P. W. (2011). Permutation methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(6), 527–542. https://doi.org/10.1002/wics.177

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100–130. https://doi.org/10.1016/j.cognition.2005.11.007

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion*, *19*(8), 1113–1139. https://doi.org/10.1080/02699930500204250

Borg, I., & Groenen, P. J. F. (2005). *Modern Multidimensional Scaling* (2nd ed., Vol. 36). Springer Science+Business Media, Inc.

Bruner II, G. C. (1990). Music, Mood, and Marketing. *Journal of Marketing, October*, 94–104.

Cohn, R., Hyer, B., Dahlhaus, C., Anderson, J., & Wilson, C. (2001). *Harmony.* Oxford University Press.

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

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

https://doi.org/10.1177/001316445601600102

Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). What music makes us

feel: At least 13 dimensions organize subjective experiences associated with

music across different cultures. *Proceedings of the National Academy of Sciences

of the United States of America, 117*(4), 1924–1934.

https://doi.org/10.1073/pnas.1910704117

Darrow, A. A., Haack, P., & Kuribayashi, F. (1987). Descriptors and Preferences for

Eastern and Western Musics by Japanese and American Nonmusic Majors.

Journal of Research in Music Education, 35(4), 237–248.

https://doi.org/10.2307/3345076

Dowling, W. J. (1978). Scale and Contour: Two Components of a Theory of

Memory for Melodies. *Psychological Review, 85*(4), 341–354.

https://doi.org/10.1037/0033-295X.85.4.341

Escofier-Cordier, B. (1965). *L'analyse des correspondances* [Doctoral Thesis].

Université de Rennes.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici,

A. D., & Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in

Music. *Current Biology, 19*(7), 573–576.

https://doi.org/10.1016/j.cub.2009.02.058

Gray, P. H., & Wheeler, G. E. (1967). The Semantic Differential as an Instrument

to Examine the Recent Folksong Movement. *Journal of Social Psychology, 72*(2),

241–247. https://doi.org/10.1080/00224545.1967.9922321

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

Academic Press.

Gregory, A. H., & Varney, N. (1996). Cross-cultural comparisons in the affective
response to music. *Psychology of Music*, *24*(1), 47–52.
https://doi.org/10.1177/0305735696241005

Hesterberg, T. (2011). Bootstrap. *Wiley Interdisciplinary Reviews: Computational
Statistics*, *3*(6), 497–526. https://doi.org/10.1002/wics.182

Juslin, P. N., & Sloboda, J. A. (Eds.). (2010). *Handbook of music and emotion:
Theory, research, applications.* Oxford University Press.

Juslin, P. N., & Västfjäll, D. (2008). All emotions are not created equal: Reaching
beyond the traditional disputes. *Behavioral and Brain Sciences*, *31*, 559–621.
https://doi.org/doi:10.1017/S0140525X08005554%20Patrik

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students.
Journal of Abnormal and Social Psychology, *28*(3), 280–290.
https://doi.org/10.1037/h0074049

Kennedy, J., Kennedy, M., & Rutherford-Johnson, T. (2013). *Harmony* (6th ed.).
Oxford University Press.

Kopacz, M. (2005). Personality and music preferences: The influence of personality
traits on preferences regarding musical elements. *Journal of Music Therapy*,
42(3), 216–239. https://doi.org/10.1093/jmt/42.3.216

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

Ladinig, O., & Glenn Schellenberg, E. (2012). Liking unfamiliar music: Effects of

felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(2), 146–154. https://doi.org/10.1037/a0024671

Madsen, C. K. (1997). Emotional Response to Music as Measured by the Two-Dimensional CRDI. *Journal of Music Therapy*, *34*(3), 187–199. https://doi.org/10.1093/jmt/34.3.187

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

Osgood, C. E., & Suci, G. J. (1955). Factor analysis of meaning. *Journal of Experimental Psychology*, *50*(5), 325–338. https://doi.org/10.1037/h0043965

Pielou, E. C. (1984). *The Interpretation of Ecological Data: A Primer on Classification and Ordination.* Wiley.

Raman, R., & Dowling, W. J. (2016). Real-Time Probing of Modulations in South Indian Classical (Carnātic) Music by Indian and Western Musicians. *Music Perception*, *33*(3), 367–393. https://doi.org/10.1525/MP.2016.33.03.367

Raman, R., & Dowling, W. J. (2017). Perception of modulations in south indian classical (carnatic) music by student and teacher musicians: A cross-cultural study. *Music Perception*, *34*(4), 424–437.

Reilly, J., Westbury, C., Kean, J., & Peelle, J. E. (2012). Arbitrary symbolism in natural language revisited: When word forms carry meaning. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0042286

Rodà, A., Canazza, S., & De Poli, G. (2014). Clustering affective qualities of classical music: Beyond the valence-arousal plane. *IEEE Transactions on*

Affective Computing, 5(4), 364–376.
https://doi.org/10.1109/TAFFC.2014.2343222

Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour, 4*(10), 1029–1038. https://doi.org/10.1038/s41562-020-0924-8

Thompson, W. F. (1994). Sensitivity to combinations of musical parameters: Pitch with duration, and pitch pattern with durational pattern. *Perception & Psychophysics, 56*(3), 363–374. https://doi.org/10.3758/BF03209770

Tucker, L. R. (1958). An inter-battery method of factor analysis. *Psychometrika, 23*(2), 111–136. https://doi.org/10.1007/BF02289009

Valentin, D., Chollet, S., Lelièvre, M., & Abdi, H. (2012). Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science & Technology*, 1–16.
https://doi.org/10.1111/j.1365-2621.2012.03022.x

Wallmark, Z. (2019). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music, 47*(4), 585–605.
https://doi.org/10.1177/0305735618768102

Wedin, L. (1969). Dimension Analysis of Emotional Expression in Music. *Swedish Journal of Musicology, 51*, 119–140.

Wedin, L. (1972). Evaluation of a Three-Dimensional Model of Emotional Expression in Music. *The Psychological Laboratories, 54*(349), 1–17.

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2014). An Interlanguage Study of Musical Timbre Semantic Dimensions and Their Acoustic Correlates. *Music*

Perception: An Interdisciplinary Journal, 31(4), 339–358.
https://doi.org/10.1525/MP.2014.31.4.339

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2015). An Interlanguage Unification of Musical Timbre: Bridging Semantic, Perceptual, and Acoustic Dimensions. *Music Perception: An Interdisciplinary Journal, 32*(4), 394–412. https://doi.org/10.1525/MP.2015.32.4.394