

Identifying Style-Types in a Sample of Musical Improvisations Using Dimensional Reduction and Cluster Analysis

Blair K. Ellis
McMaster University

Heungsun Hwang
McGill University

Patrick E. Savage
Tokyo University of the Arts

Bing-Yi Pan and Annabel J. Cohen
University of Prince Edward Island

Steven Brown
McMaster University

Creativity research examines both the processes and products of creativity. An important avenue for analyzing creativity is by means of spontaneous improvisation, although there are major challenges to characterizing the products of improvisation because of their variable nature. A useful concept missing from the analysis of improvisation is the idea that the products of a corpus can be organized into a series of “style-types,” where each type differs from others in certain key structural features. Clustering methods provide a reliable quantitative means of examining the organization of style-types within a diverse corpus of improvisations. To look at the utility of such methods, we examined a sample of 72 vocal melodic improvisations produced by novice improvisers. We first classified the melodies acoustically using a multidimensional musical-classification scheme, which coded the melodies for 19 distinct features of musical structure. We next employed multiple correspondence analysis (a dimensional reduction method) and k-means cluster analysis simultaneously, and obtained 3 relatively discrete clusters of improvisations. Stylistic analysis of these clusters revealed that they differed in key musical features related to phrase structure and rhythm. Cluster analyses provide a promising means of describing and analyzing the products of creativity, including variable structures like spontaneous improvisations.

Keywords: vocal improvisation, AIRS Test Battery, singing, musical classification, style analysis

Supplemental materials: <http://dx.doi.org/10.1037/aca0000072.supp>

Creativity is not just a process of generating novelty, but a process of generating diversity. Creative work results in a more diverse set of products than there was to begin with. For example, in modern times, new cell phone “apps” are being created on a regular basis, diversifying the potential uses of cell phones. The generation of novel products can be examined either along the broad time scales required to produce apps (Joorabchi, Mesbah, & Kruchten, 2013), musical compositions (Collins, 2007), and sci-

entific theories (Finke, Ward, & Smith, 1992; Mace & Ward, 2002), or along the shorter time scales involved in spontaneous creativity, such as musical improvisation (Beaty, 2015). Work on spontaneous creativity has followed two major streams, one focused on conceptual processing (e.g., divergent thinking tasks, brainstorming) and the other on performance.

For performance, musical improvisation has provided a rich set of findings and theories that have shed light on the underlying

This article was published Online First December 19, 2016.

Blair K. Ellis, Department of Psychology, Neuroscience, & Behaviour, McMaster University; Heungsun Hwang, Department of Psychology, McGill University; Patrick E. Savage, Department of Musicology, Tokyo University of the Arts; Bing-Yi Pan and Annabel J. Cohen, Department of Psychology, University of Prince Edward Island; Steven Brown, Department of Psychology, Neuroscience, & Behaviour, McMaster University.

Blair K. Ellis is now at the Department of Health and Rehabilitation Sciences, Western University. Bing-Yi Pan is now at the Department of Creative Arts Therapies, Concordia University.

This work was funded in part by a grant from the Natural Sciences and Engineering Research Council of Canada to Steven Brown (grant 371336) and by a Social Sciences and Humanities Research Council of Canada Major Collaborative Research Initiative supporting the Advancing Interdisciplinary Research in Singing project (grant 412-2009-1008, Director Annabel J. Cohen) providing summer internships to Blair K. Ellis and postdoctoral support to Bing-Yi Pan.

Correspondence concerning this article should be addressed to Steven Brown, Department of Psychology, Neuroscience, & Behaviour, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4M9, Canada. E-mail: stebro@mcmaster.ca

processes of spontaneous creativity (e.g., Norgaard, 2011, 2014). Unlike work on divergent thinking (Silvia et al., 2008), research on improvisation has typically looked at specialists, notably jazz instrumentalists and singers, with a lifetime of training in performance and a professional training in improvisation itself. Not only do such people have a deep personal investment in creative work, but they are able to verbally describe their individual approaches to improvisation (Biasutti & Frezza, 2009; Norgaard, 2011; Wopereis, Stoyanov, Kirschner, & Van Merriënboer, 2013). Theoretical models of improvisation have tended to emphasize the procedural aspect of its mechanisms. According to Pressing's (1988, 1998) model—which is a well accepted theory for jazz improvisation—improvisers insert prelearned musical structures at stylistically appropriate times, and use perceptual feedback and error correction to monitor production and make adjustments between intention and actualization. Improvisers may also choose to incorporate material that is completely unrelated to the present context (Norgaard, 2014), further diversifying the product.

The study of musical performance has revealed both the promise and the challenges of studying improvisation. Studies using a cognitive-scientific approach have attempted to shed light on the mechanisms of improvisation by examining the effects of experimental manipulations on production, such as improvising in familiar versus unfamiliar keys (Goldman, 2013) or improvising during a dual-task paradigm (Fidlon, 2011). Others have analyzed the relationship between cognitive factors like divergent thinking (Beaty, Smeekens, Silvia, Hodges, & Kane, 2013) or working memory (De Dreu, Nijstad, Baas, Wolsink, & Roskes, 2012) and expert ratings of perceived creativity. Although these experimental approaches show promise in elucidating the cognitive mechanisms of improvisation, they also provide analytical challenges in characterizing the multidimensional nature of the improvised *products* themselves. Examples of dependent measures used to represent musical structure in these studies have included note entropy and the proportion of diatonic pitch classes (Goldman, 2013), the number of repeated intervals and rhythmic patterns (Norgaard, 2014), and the statistical distribution of tone choices across metrically salient beats (Järvinen, 1995). Although studies like these represent important approaches to examining musical improvisation in their own right, few attempts have been made to analyze improvisational products in a more holistic and multivariate manner, most especially with methods that do not require transcription (see Hickey & Lipscomb, 2006; Madura Ward-Steinman, 2008; Raju & Ross, 2012).

This present study aims to address the need for a quantitative and multidimensional approach for the analysis of the behavioral products of improvisation that simultaneously takes into account the multiple structural musical features that performers are able to manipulate while improvising. The approach takes advantage of methods coming from classification theory and cluster analysis. In particular, our approach begins by classifying a large sample of musical improvisations using the musical classification scheme “CantoCore” (Savage, Merritt, Rzeszutek, & Brown, 2012) to understand the diversity of musical structures in the sample. CantoCore is a multidimensional coding scheme that contains 26 characters of musical structure that span the broad domains of rhythm, pitch, syllable, texture, and form. Each of the 26 characters contains 3 to 6 character-states. This scheme was previously used to classify a highly complex sample of 259 traditional group-

level vocal songs derived from 12 indigenous populations of Taiwan (Rzeszutek, Savage, & Brown, 2012; Savage & Brown, 2014). The observed clusters differed from one another in a multidimensional manner; in other words, the clusters represented different conglomerations of musical features, and thus could be thought of as stylistic song-types. The term “cantogroup” (where the root “canto” means song) was developed to describe these stylistic song-types (Savage & Brown, 2014).

The goal of the current study was to apply a style analysis to a sample of musical improvisations after coding them with CantoCore. In thinking about the concept of musical style, we make reference to Leonard Meyer's classic book *Style and Music* (Meyer, 1989), in which he described classification and style analysis in the following way:

Classification is essentially a descriptive discipline. It tells us what traits go together and with what frequency they occur, but not why they do so. Style analysis is more ambitious. It seeks to formulate and test hypotheses explaining why the traits found to be characteristic of some repertory—its replicated melodic patterns, rhythmic groupings, harmonic progressions, textures, timbres, and so on—fit together, complementing one another. (p. 43)

To achieve our goal, we used an extension of multiple correspondence analysis (MCA; Abdi & Valentin, 2007) involving the simultaneous implementation of k-means cluster analysis (Hartigan & Wong, 1979). MCA is a powerful data-reduction technique (Husson & Josse, 2014) that can be used to characterize individual objects (e.g., improvisations) based on the variables with which they are associated (e.g., structural musical features). k-means cluster analysis is an unsupervised clustering algorithm that, for our purposes, can be used to divide a sample of musical improvisations across a reduced number of MCA-derived dimensions into a set of relatively discrete style-clusters. Rather than characterizing individual improvisations based on the variables that they are highly associated with, as would occur in traditional MCA, this method is designed to characterize discrete subsets (clusters) of similar improvisations based on the variables that best define each subset (Hwang, Dillon, & Takane, 2006). The final result is an integrated graphical display (i.e., a 2D or 3D plot) that permits easy interpretation of relatively discrete clusters of improvisation-styles within the sample, as well as the interrelationships among these styles based on the musical features on which they were coded.

The principal objective of the current study was to examine whether we could identify subsets of style-types within a corpus of spontaneous vocal improvisations, as well as provide insights into the different manners of improvising that underlie these various style-types. We analyzed a set of 72 vocal melodic improvisations generated as part of the AIRS (Advancing Interdisciplinary Research in Singing) Test Battery of Singing Skills (ATBSS; Cohen, 2015; Cohen, Armstrong, Lannan, & Coady, 2009), in which the participants were novice improvisers. We coded all of the improvisations acoustically in a multidimensional manner using a modified version of the CantoCore classification scheme designed for monophonic melodies, and then subjected the data to a simultaneous MCA and k-means cluster analysis to generate relatively discrete clusters of improvisational style. The goal then was to use the multivariate plot resulting from the analysis to ascertain the key stylistic differences among the clusters, as based on their

CantoCore codings. From this, it should be possible to make mechanistic inferences about the improvisation process based on which features are most stable across the improvisations and which features are most variable. In addition to doing a structural analysis of the improvisations, we had a group of professional musicians rate the improvisations for their levels of creativity and performance quality to look for relationships between the musical style (as defined by the structural musical features) and external assessments of creativity and performance quality.

Method

The Improvisation Sample

The vocal improvisation data were collected using the online interactive AIRS Test System (Pan & Cohen, 2012, 2016). This system is an automated, audiovisual, computer-based version of the original in-person, interview-based version of the ATBSS (Cohen et al., 2009), which assesses 11 main components of human singing skills. The test was conducted in a lab setting, as described in the Procedure section below. Data for the improvisation task of ATBSS were analyzed for the current study (described in detail below). Audiovisual data from the test battery were uploaded to an online database, which was accessible by researchers affiliated with the AIRS project.

Participants. At the time that the study began, 110 adult participants (67 females, 43 males; age: $M = 31.7$, $SD = 17.3$) had taken part in the ATBSS. They were recruited through posters and word of mouth. They completed the test battery, followed by an online e-mail survey created at the University of Prince Edward Island (UPEI) about their linguistic and musical backgrounds, including a 10-point test assessing knowledge of music theory and music reading skills (unpublished). A participant's score on this 10-point test will be referred to as a "music test score" throughout this paper. From these 110 original participants, data from 38 were excluded. Reasons for exclusion included: instances where participants generated improvisations that copied preexisting music presented earlier in the test battery or well-known songs that existed in pop culture; our inability to code the samples due to the indecisiveness of the singer during improvising (i.e., stopping and starting); samples that were spoken rather than sung or that were incomplete; or an absence of samples due to technical difficulties with the recordings. Other reasons for exclusion included participants who sang both melodies with or without words (rather than one with words and one without; see Procedure below), and one participant who was not a native English speaker and who did not sing in English during the improvisation with words. The final number of participants was 72 (44 females, 28 males; musician-ship: 33 musicians, 39 nonmusicians, as based on self-ratings; age: $M = 30.8$, $SD = 16.7$). The study was approved by the Research Ethics Board of UPEI.

Procedure. Participants were tested in a double-walled sound-attenuated chamber. The test was presented to participants using the AIRS Test System (Pan & Cohen, 2012) on a Mac Pro computer that was connected to a 19-inch LCD monitor and two PSD Synchrony One B Speakers. Audiovisual responses were recorded (AIRS Test System, 44.1 kHz, 16 bits, 15fps) with a Blue Microphone EyeBall 2.0 HD webcam, and were saved remotely in

the ATBSS database. The duration of the entire 11-component test battery was around 30 min.

Data from the improvisation task of the ATBSS were analyzed for the current study. During this task, participants sat in front of a computer monitor, were presented with a set of 4 images (heart, flower, sun, and apple), and were asked to select one. After doing so, the 3 unselected images disappeared from view, and the participant was asked to mentally create a song inspired by the selected picture. The images were used to facilitate engagement in this unusual request to generate a melody on the spot. Participants were then asked to click on an audiovisual "record" button and sing their improvisation when they felt ready. The same procedure was repeated for a second improvisation using a selection of one of the three remaining images. Of the two melodies generated by each participant, one contained words and the other did not. For the latter, participants used a self-selected vocable like "la." There was no instruction about whether or not to use words in the first improvisation. For the second improvisation, participants were instructed to sing with words if their first melody was sung without words, and vice versa. This resulted in a total of 144 improvisations generated by the 72 participants. We decided to restrict our analysis to the samples without words because of our general interest in musical creativity, rather than language creativity. Hence, the final dataset consisted of the 72 improvisations without words (i.e., one improvisation per participant).

Ratings of Creativity and Performance

Participants. Six university professors (4 male, 2 female; age: $M = 56$, $SD = 11$) at a Canadian university's department of Music participated in the rating experiment, all of whom held either an MA, DMA, or PhD in music performance or music theory. They were recruited via e-mail invitation. They completed a demographic questionnaire that included items about their musical experiences, listening, and training. All participants reported an absence of hearing problems that might have influenced their music listening.

Procedure. The six participants were pseudorandomly assigned to either a creativity-rating condition ($n = 3$) or a performance-rating condition ($n = 3$). After the participant signed a consent form, the experimenter (Blair K. Ellis) read general instructions about the experimental procedure to each of the six participants individually. Each participant sat in his or her university office for the duration of the test, which took approximately one hour. A Macbook Pro computer, a set of MA-10ABK Ediol speakers ($9'' \times 6'' \times 7''$), and the presentation software PsychoPy were used to present participants with text-based instructions that allowed them to navigate through the PsychoPy program.

The program first presented 10 randomly selected audio recordings of improvisations from the ATBSS that were not used in the rating experiment. These melodies were presented for listening purposes to give the raters a sense of what the corpus of improvisations sounded like as a whole. This was done to establish the raters' expectations of the overall musical level of the samples and the musical skill of the participants. The raters were then prompted with a message informing them that they would be asked to listen to and then rate the improvisation samples ($n = 144$) one at a time in a randomized order on a scale of either creativity or performance quality, depending on the condition that the rater was in.

Although the full set of 144 improvisations was rated (72 without words and 72 with words), only the ratings for the samples without words are reported here for the reasons mentioned above.

The “consensual assessment technique” (CAT; Amabile, 1982) is considered to be the gold standard for assessing creative products (Carson, 2006; Kaufman, Baer, & Cole, 2009). It assumes that experts within a given domain should agree on the creative assessment of a sample. Hence, the test is not associated with any particular theory of creativity (Baer & McKool, 2009). Raters in the creativity condition were asked to use their own definition when making ratings of creativity. However, we clarified that their ratings should be based on *structural* features only (e.g., melody, rhythm, phrase structure, musical form), and that performance features (performance quality, performance competence of the singer, quality of the voice, etc.) should not influence their judgment. Raters in the performance-rating condition were instructed to do precisely the opposite and focus on performance features, but not structural features or creativity.

Classifying the Improvisation Samples

The audiovisual improvisation samples were downloaded from the online test-battery database and were converted into audio-only files using a script in the Apple terminal. To classify the improvisations, we used a modified version of the musical-classification scheme CantoCore (Savage et al., 2012). The modified version consisted of 19 “characters” of musical structure for solo vocal improvisation. These are outlined in Table 1, along with the labels used for each of the “character-states,” or levels of the characters (e.g., character 10 = number of pitch classes, where the character-states = 101 [few pitches], 102 [moderate number of pitches], 103 [many pitches]) that are used for the plot in Figure 3. When the 19

musical characters are expanded into their respective levels, the result is 68 possible character-states (see Table S1 in the supplementary materials for the complete coding scheme, including the definitions of all character-states). The coder (Blair K. Ellis) listened to each of the melodies and coded them according to the scheme in Tables 1 and S1. For a given sample, the coder selected a single character-state for each one of the 19 characters, which resulted in a data matrix where each row corresponded to a unique musical profile across all 19 characters of musical structure for a single improvisation (see Figure S1). Items that were not codable or that were arbitrary were marked as “NA” for not applicable and were treated as missing data. As recommended by Savage et al. (2012), the coder listened to each musical sample as often as necessary to arrive at an accurate coding. We will also refer to the character-states as “musical features” throughout the paper.

We modified the published coding scheme to adapt it to solo vocal improvisations (Table S1). All characters dealing with musical texture (i.e., multiple musical lines, as in the case of polyphony) were omitted, because all of the improvisations were solo vocal melodies. The “melisma” character was omitted because all samples were sung without words. Samples were sung using various syllables, where melismas often sounded arbitrary or were difficult to detect. To quantify the length of each improvisation, a new character called “total number of phrases” (line 15) was added to the scheme. For this character, we counted the total number of musical phrases in the sample, rather than measuring the sample’s duration in seconds or minutes. In addition, we added a new character that measured “repetition with variation” (line 17). This character was captured by coding the proportion of phrases that contained repetition (i = < 25% of phrases, ii = 25% to 50%, iii = > 50%). The rationale for including this character was that

Table 1
Modified Coding Scheme Used to Code the Vocal Improvisations

Number	Character	Musical structure	Character-state	Graphical label
1	Meter	Rhythm	a, b, c, d	01a, 01b, etc.
2	Number of beats	Rhythm	a, b, c, d	02a, 02b, etc.
3	Beat sub-division	Rhythm	a, b, c, d	03a, 03b, etc.
4	Number of sub-beats	Rhythm	a, b, c, d	04a, 04b, etc.
5	Syncopation	Rhythm	i, ii, iii	51, 52, 53
6	Motivic redundancy	Rhythm	i, ii, iii	61, 62, 63
7	Durational variability	Rhythm	i, ii, iii	71, 72, 73
8	Tonality	Pitch	a, b, c, d, e	08a, 08b, etc.
9	Mode	Pitch	a, b, c, d, e	09a, 09b, etc.
10	Number of pitch classes	Pitch	i, ii, iii	101, 102, 103
11	Hemitonicity	Pitch	i, ii, iii	111, 112, 113
12	Melodic interval size	Pitch	i, ii, iii	121, 122, 123
13	Melodic range	Pitch	i, ii, iii	131, 132, 133
14	Melodic contour	Pitch	a, b, c, d, e, f	141, 142, 143
15	Total number of phrases	Form	i, ii, iii	151, 152, 153
16	Phrase repetition	Form	i, ii, iii	161, 162, 163
17	Repetition with variation	Form	i, ii, iii	171, 172, 173
18	Phrase length	Form	i, ii, iii	181, 182, 183
19	Phrase symmetry	Form	i, ii, iii	191, 192, 193

Note. The coder listened to the samples and assigned character-states for each of the 19 characters of the modified CantoCore scheme (see Table S1 for definitions of each character-state). The letter options represent categorical (nominal) options, while the number options represent ordinal options (i = *little or none*, ii = *moderate*, iii = *high*, all treated as categorical in this paper for the purposes of MCA). Graphical labels are presented for the purposes of interpreting the musical character-states in Figure 3.

an improvisation that included repetition with variation introduced an element of complexity that was not codable in the original scheme, namely variation, in addition to overall repetition per se. It is important to note that repetition with variation is a nested character within repetition (line 16), because it is not possible to have varied repetition without having repetition itself.

To measure interrater reliability, a second rater coded a randomly selected portion (20%) of the samples. The rater was trained to use CantoCore and was given instructions regarding the modifications made to the coding scheme. She was a graduate from a Bachelor of Music program, and had a similar level of music training as the principal coder. Interrater reliability was calculated for each individual character of the modified CantoCore coding scheme in two ways. One measure was the percent agreement between the two raters. In addition, Cohen's Kappa was calculated, which is thought to better account for the effects of chance agreement, partial agreement, and character redundancy (Savage et al., 2012). The overall percent agreement between the two raters was 62.4%, while Cohen's Kappa was 0.32. According to Landis and Koch (1977), Cohen's Kappa values of 0.21–0.40 are considered “fair” agreement. Note that these values are arbitrary, and that the use of Cohen's Kappa for interrater agreement has been criticized (e.g., for its dependence on marginal distributions; Banerjee, Cappozzoli, McSweeney, & Sinha, 1999). However, the percent agreement and Cohen's Kappa values reported here are similar to those observed in previously published studies using CantoCore (Brown et al., 2014; Savage, Brown, Sakai, & Currie, 2015; Savage et al., 2012).

Cluster Analysis of the Improvisation Samples

The main goal of this analysis was to use clustering techniques to identify relatively discrete clusters of improvisations that represent different manners of creating spontaneous musical improvisations in terms of structural musical features. The method of analysis is outlined in Hwang et al. (2006). A simplified descrip-

tion of the data-reduction and clustering procedure is outlined in Table 2 (steps 2–4). Using this approach, MCA and k-means cluster analysis were applied simultaneously in a single framework to (a) identify a low-dimensional feature space for music that represents the different character-states of the 19 musical characters of the coding scheme, and (b) identify and describe the musical structures of relatively homogeneous improvisation clusters within this low-dimensional feature space (Hwang et al., 2006). If successful, the improvisations within the resulting clusters should have similar codings based on the modified CantoCore coding scheme. In other words, the clusters should represent distinct improvisation styles, as based on musical structure.

The raw CantoCore-coded dataset was a matrix that consisted of 72 rows (one for each improvisation) and 19 columns (one for each of the CantoCore characters). The raw codings were then transformed into an “indicator matrix”, where each character-state was represented as present (1) or absent (0; see Figure S1). This resulted in transforming the 72×19 matrix into a 72×68 matrix because the musical characters were expanded into their respective character-states and coded for their presence versus absence (see Table S1 in the supplementary materials for the complete coding scheme, including the definitions of all character-states). Finally, preprocessing of the 72×68 matrix was necessary to ensure a stable MCA solution. This consisted of removing all features in which 3 or fewer samples from the entire set of 72 improvisations were coded as “present” for that feature. This removed 22 of the 68 features, resulting in a final indicator matrix of 72 improvisations (rows) by 46 music musical features (columns).

MCA was conducted on the final indicator matrix to determine the number of reduced dimensions. A three-dimensional solution was selected (i.e., $d = 3$, where d refers to the number of dimensions) because the size of the “adjusted inertia”—which corresponds to a measurement of variance in MCA—decreased more slowly after the first three values (see Figure

Table 2
Summary of the Analysis Method for Dimensional Reduction and Cluster Analysis

Step	Order	Level of analysis	Method	Description
1. Classify the improvisations	1	Pre-data analysis, considering songs independently	CantoCore	Classify each improvisation on 19 musical characters based on listening
2. Select the number of dimensions	2	Pre-data analysis, all improvisations in the sample are analyzed together	Data reduction using MCA	Select the number of dimensions based on maximizing the variance explained and minimizing the number of dimensions (Figure S2)
3. Select the number of clusters	3	Pre-data analysis, all improvisations in the sample are analyzed together	Under the fixed number of dimensions, conduct the simultaneous MCA and k-means analysis iteratively with increasing numbers of clusters	Select the number of improvisation clusters based on minimizing the optimization criterion and minimizing the number of overall clusters (Figure 1)
4A. Multiple correspondence analysis (MCA)	4A and 4B occur simultaneously	Data analysis, all improvisations in the sample are analyzed together	Data reduction using MCA	Uncover a low-dimensional multivariate musical feature space that explains variance in the data
4B. K-means cluster analysis			K-means cluster analysis is conducted on improvisation samples with reference to MCA results	Assign improvisations to clusters based on stylistic similarities

S2). Adjusted inertia values are a better approximation of the variance in MCA than are unadjusted inertias (Abdi & Valentin, 2007; Benzécri, 1979). The three-dimensional solution accounted for 69.15% of the total adjusted inertia (variance) in the data ($D1 = 31.45\%$, $D2 = 24.12\%$, and $D3 = 13.58\%$). Given the predetermined number of dimensions, the simultaneous approach to MCA and k-means clustering was applied to the indicator matrix, varying the number of clusters. The number of clusters was selected by examining the scree plot, and ensuring the parsimonious balance between minimizing the optimization criterion and the overall number of clusters (see Figure 1). Minimizing the optimization criterion is equivalent to finding a low-dimensional representation for the numerous categorical musical variables while simultaneously identifying clusters of improvisations in the sample that are relatively homogenous in terms of these musical variables (Hwang et al., 2006). The number of clusters was selected to be 3 (i.e., $c = 3$, where c refers to the number of clusters), as defined by the “elbow” (i.e., transition point) of scree plot in Figure 1 (Everitt & Horton, 2014). The simultaneous MCA and cluster analysis was conducted using MATLAB (MathWorks Inc., R2014b).

Rating Analysis

Three musical experts rated each improvisation on a Likert scale from 1 to 7 for musical creativity, and three different experts rated each sample on a Likert scale from 1 to 7 for performance quality. Six Pearson product–moment correlations were calculated to determine interrater reliability between pairs of raters for creativity and between raters for performance (Table S2). Creativity ratings for Rater 3 were discarded due to a strong discrepancy between their ratings and those of creativity Raters 1 and 2. Despite the proposed power of the CAT (Amabile, 1982), this rating discrepancy is understandable. The background of Rater 3 was primarily as a high-level music performer and music performance educator, whereas the other two creativity raters had substantially greater experience as music theorists, which could explain different strategies for

rating creativity. Given the high interrater consistency in creativity ratings for Raters 1 and 2, these ratings were scaled, centered, and averaged separately across the two raters so that each improvisation had a single composite rating for creativity. The same was done for the performance ratings, except that the single composite rating for performance was based on all three raters, due to higher interrater consistency.

To test our hypotheses that there would be one or more mean differences between (a) cluster membership and creativity ratings, and (b) cluster membership and performance ratings, two univariate linear models were performed. Each model included the same covariate (music test score), factor (cluster membership), interaction between the covariate and factor, and predicted either (a) creativity ratings or (b) performance ratings. Preliminary data screening procedures were conducted prior to each analysis to ensure that the assumptions for univariate tests were met. Although these data may seem appropriate for a multivariate linear model, the decision to perform two separate univariate models was made with reference to our research questions. Because univariate and multivariate linear models inherently aim to resolve different research questions, it is important that care is taken for the selection of analyses that are appropriate to research questions being asked (Huberty & Morris, 1989). In our case, we aimed to examine our dependent measures separately because the linear combination of both variables together was not of interest. Statistical analyses for the rating data were conducted using R statistical software 3.2.3 (R Core Team, 2015).

Results

Simultaneous MCA and K-means Cluster Analysis

Standard results for MCA analyses consist of two sets of coordinates in Euclidean space. Using the terminology of MCA (Greenacre, 2007), these are referred to as *individual coordinates* (i.e., the improvisations) and *variable coordinates* (i.e., the coded musical features). Individual coordinates that are close to one another in space are similar, whereas individual coordinates that are far apart are dissimilar, and the same is true for the variable coordinates. The interpretation of the results occurs by examining associations between the two coordinate-sets after superimposing them in a single graphical display (see Figure 3). When particular individual coordinates appear close to (rather than far from) variable coordinates, the two coordinate sets tend to be highly related, such that the individual coordinate can be characterized by the variable coordinate (Abdi & Valentin, 2007).

Figure 2 demonstrates the manner by which the individual improvisations (as shown by unfilled circles in the plot) were grouped into three relatively discrete clusters of improvisation styles: C1, $n = 30$; C2, $n = 32$; C3, $n = 10$. The black circles represent the centroid of each cluster. The improvisations cluster strongly with reference to MCA dimensions 1 and 2, but not to dimension 3. It is important to note that interpreting dimensions as continua is not straightforward in this analysis because of rotational freedom. Thus, it is more reasonable to focus on how the improvisations, musical features, and/or cluster centroids are located relative to one another in Figure 3.

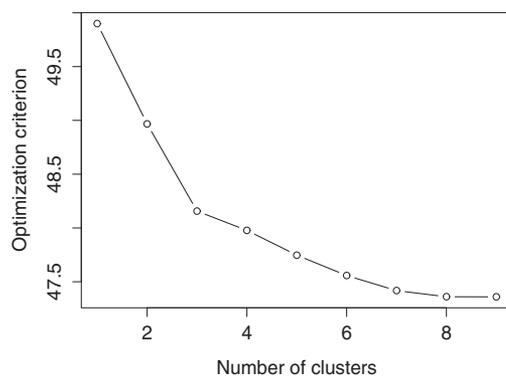


Figure 1. Scree plot for the number of clusters. The number of clusters can be selected by examining the scree plot and ensuring the parsimonious balance between minimizing the optimization criterion and the overall number of clusters (Hwang et al., 2006). The number of clusters for this MCA solution was 3, as defined by the “elbow” (i.e., transition point) of the scree plot (Everitt & Horton, 2014).

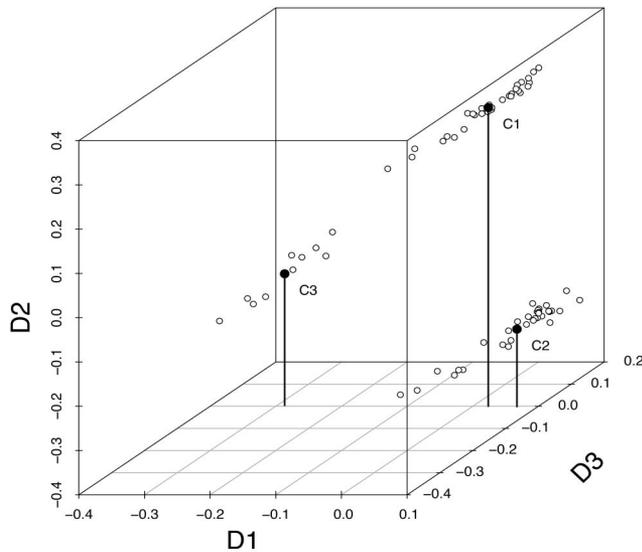


Figure 2. Dimensional plot of the individual coordinates (i.e., improvisations) and their cluster membership. The unfilled circles represent the individual improvisations, and the black circles in the center of each cluster represent the cluster centroid for each of the three improvisation styles. D1, D2, and D3 refer to the first three dimensions of the MCA analysis. Dimension 3 does not contribute strongly to cluster membership, as the coordinates for each of improvisations are spread out along this dimension.

Figure 3 represents the same improvisation clusters as in Figure 2, but in reference to the variable coordinates, namely the 46 character-states of the 19 characters of CantoCore on which the improvisations were coded (see Table S1 for a definition of the graphical labels shown in the plot). Because the clusters are distributed evenly along the third dimension, as shown in Figure 2, we will focus here on the plot of dimensions 1 and 2. As commonly occurs with cluster analyses, the results in Figure 3 do not demonstrate a nonoverlapping solution. This suggests that the dataset as a whole is rather homogeneous in terms of musical features, and that the clusters differ from one another on a small number of musical features.

Interpreting Improvisational Styles From Cluster Membership

Cluster 1. Based on Figure 3, we can see that the central upper portion of the plot is related to the first cluster, C1. Its cluster centroid seems to be more strongly associated with musical features “133” (large melodic range) and “161” (nonrepetitive phrases) than any other features. The improvisations in C1 ($n = 30$) make up 41.7% of the sample. Panel A of Figure S3 in the supplementary materials represents the relative percentages of presence for each of the 46 musical features for all improvisations in C1. Based both on these results and on listening to the samples, we find that C1 contains improvisations that are typical of the Western musical idiom. They are primarily major-key, iso-tonal improvisations, and they tend to be in simple common time (2/4, 4/4). The phrases are mostly through-composed and thus nonrepetitive.

Cluster 2. The bottom right portion of Figure 3 is associated with the second cluster, C2. The cluster centroid seems to be more strongly associated with musical features “173” (high degree of repetition with variation), “191” (high degree of phrase symmetry), “131” (small melodic range), and “151” (small number of overall phrases). The improvisations in C2 ($n = 32$) make up 44.4% of the overall sample. Panel B of Figure S3 represents the relative percentage of presence for each of the 46 musical features in all improvisations in C2. Based both on these results and on listening to the samples, we find that C2, like C1, contains improvisations that sound typical of the Western musical tradition. They are primarily major-key, iso-tonal improvisations, and they tend to be in simple common time. In contrast to the samples in C1, there is a very high degree of phrase repetition and repetition with variation in C2. These improvisations seem to demonstrate highly symmetrical phrases and a more organized compositional structure, as compared with samples in the other two clusters.

Cluster 3. The bottom left portion of Figure 3 is related to the third cluster, C3. The cluster centroid seems to be most strongly associated with musical feature “01b” (heterometric meter, i.e., semiregular rhythmic patterns that contain multiple successive meters), although this is difficult to interpret as it is not nearly as close to the cluster centroid as is the case for the variable/centroid relationships for C1 and C2. 13.9% of the improvisations were grouped in C3 ($n = 10$). Panel C of Figure S3 represents the relative percentage of presence for each of the 46 musical features in all improvisations in C3. Based both on these results and on

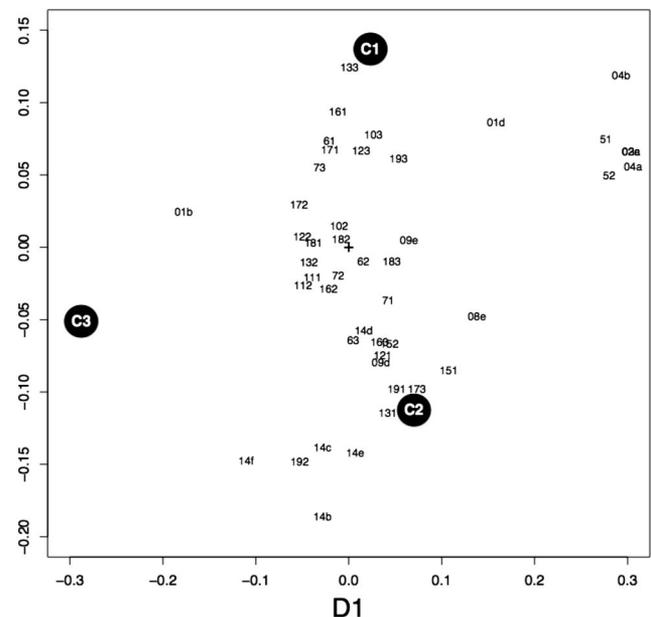


Figure 3. Dimensional plot of the variable coordinates (i.e., musical features) with the centroids of each improvisation cluster superimposed. The centroid of each cluster is represented by a black circle with the cluster number in the center. “Variable coordinates” represent the character-states of the 19 CantoCore musical characters (see Table S1 for label definitions of the character-states). The variable coordinates that are located nearest to a given cluster centroid indicate a strong association between the variable coordinate and a particular cluster of improvisations. D1 and D2 refer to the first two dimensions of the MCA analysis.

listening to the samples, we find that C3 seems to be primarily defined by heterometric rhythms or by pauses, hesitations, or unusual phrase combinations that made meter difficult to interpret.

Between-Cluster Differences

The fact that stylistic heterogeneity between the clusters can be explained by only 7 of the 46 musical features (see Figure 3) suggests that the sample is relatively homogenous overall. In other words, the relative distribution of the presence of these 39 features in the improvisation samples is more or less equal across the 3 clusters, suggesting that these features are not uniquely related to any single cluster. We now focus on those features that did distinguish the clusters.

Metric (C1, C2) versus hetero-metric (C3). An important distinction in the results is between improvisations that follow “typical Western musical ideals” related to rhythm and meter (e.g., mostly common duple or triple meters) and those melodies that do not (e.g., defined by heterometric rhythms, or pauses and hesitations, etc.). In the top right-hand corner of Figure 3, the musical features “01d” (isometric meter), “02a” (number of beats is duple), “03c” (iso-divisive subdivisions), “04a” (simple subbeats), and “51” (little or no syncopation) are grouped together. These features are present in C1 and C2 to a similar degree, although they are not prevalent in C3. The presence of these particular features is indicative of the fact that the former two clusters demonstrate rhythmic structures that are mostly in 4/4 or 3/4 meter. In contrast, musical feature “01b” (heterometric meter) is on the opposite (left) side of Figure 3, closer to C3. This suggests that C3 contains improvisations with rhythms that are either semiregular and that contain multiple successive meters (Savage et al., 2012) or that contain irregular hesitations that are likely due to a lack of musical experience.

Repetition/variation (C2) versus no repetition/variation (C1, C3). Another important distinction is between those improvisations that contain musical material that is repeated and that varied over the course of the melody, compared with those that do not show this repetition with variation. Musical features “163” (highly repetitive phrases) and “173” (high proportion of repetition with variation) appear to be much more related to C2 than to either C1 or C3, suggesting that C2 contains much more material with variation, rather than through-composed material, when compared with C1 and C3.

Rating Analysis

Creativity. Prior to analysis, the normality assumption was assessed for the creativity rating univariate linear model using both the Shapiro-Wilk normality test, $W = 0.98$, $p = .59$, and the Lilliefors (Kolmogorov–Smirnov) normality test, $d = 0.084$, $p = .24$, and both revealed no significant violation of the assumption of normality. Levene’s test was used to examine whether there were serious violations of the homogeneity of variance assumption across groups, but no significant violation was found, $F(2, 69) = 0.25$, $p = .78$.

A univariate linear model was used to test if there was a significant difference in creativity ratings between musical styles (cluster membership), if music test scores could significantly predict performance ratings, and if there was an interaction between

music test scores and musical style. Results showed a significant interaction between membership and music test scores, $F(2, 66) = 4.14$, $p = .020$, $\eta^2 = 0.11$, which suggests that the effect of music test score as a predictor of creativity ratings differed between clusters. Although the effect of music test score was also significant, $F(1, 66) = 8.44$, $p = .005$, $\eta^2 = 0.11$, we will focus on the significant interaction. The effect of cluster membership was not significant, $F(1, 66) = 2.37$, $p = .10$, $\eta^2 = 0.067$, which indicates that there were no significant differences between mean creativity ratings for C1 ($M = 0.027$; $SD = 0.82$), C2 ($M = 0.16$; $SD = 0.93$), and C3 ($M = -0.59$; $SD = 0.85$). Figure 4A shows a plot describing this analysis, including the 95% confidence intervals.

Given the significant interaction, three follow-up bivariate regression analyses were performed to test if music test scores significantly predicted creativity ratings within each cluster. To correct for multiple comparisons, Holm’s sequential Bonferroni test was used. For C2, results revealed a statistically significant effect of music test scores, $\beta = 0.14$, $\tau(30) = 4.33$, $p = .00016$. The r^2 for this equation was 0.38; that is, 38% percent of the variance in creativity ratings was predicted from music test scores. The correlation between music test scores and creativity ratings for improvisations in C2 was statistically significant, $r(30) = 0.62$, $p = .000082$. The effect of music test scores was not significant for C1, $\beta = -0.0023$, $\tau(28) = 0.063$, $p = .95$, $r^2 = 0.00014$, or C3, $\beta = -0.076$, $\tau(8) = 0.063$, $p = .46$, $r^2 = 0.070$. The correlations between music test scores and creativity ratings for improvisations in C1 ($r(28) = -0.012$, $p = .52$) and C3 ($r(8) = 0.26$, $p = .23$) were not significant.

Performance quality. The assumption of normality was met for the performance-rating univariate model, as evidenced by the Shapiro-Wilk normality test, $W = 0.98$, $p = .47$, and the Lilliefors (Kolmogorov–Smirnov) normality test, $D = 0.087$, $p = .19$. Levene’s test for homogeneity of variance showed no violation of the homogeneity of variance assumption $F(2, 69) = 0.085$, $p = .92$.

A univariate linear model was conducted to test whether there was a significant difference in performance ratings between musical styles (cluster membership), whether music test scores could significantly predict performance ratings, or whether there was an interaction between music test scores and musical style. Results showed no significant interaction between music test scores and cluster membership, $F(2, 66) = 0.40$, $p = .67$, $\eta^2 = 0.012$. Additionally, there was no significant main effect of cluster membership, $F(2, 66) = 1.17$, $p = .32$, $\eta^2 = 0.034$, which indicates that there were no significant differences between mean performance ratings for C1 ($M = -0.14$; $SD = 0.85$), C2 ($M = 0.12$; $SD = 0.84$), and C3 ($M = -0.37$; $SD = 0.87$). However, there was a significant main effect of music test scores, $F(1, 66) = 7.98$, $p = .006$, $\eta^2 = 0.11$, which suggests there is a significant positive relationship between music test scores and performance ratings. Figure 4B presents a plot describing this analysis, including the 95% confidence intervals.

Discussion

The principal objective of this study was to shed light on both the products and processes of musical improvisation. We did this by applying classification and clustering methods to a sample of vocal melodic improvisations produced by novice improvisers in

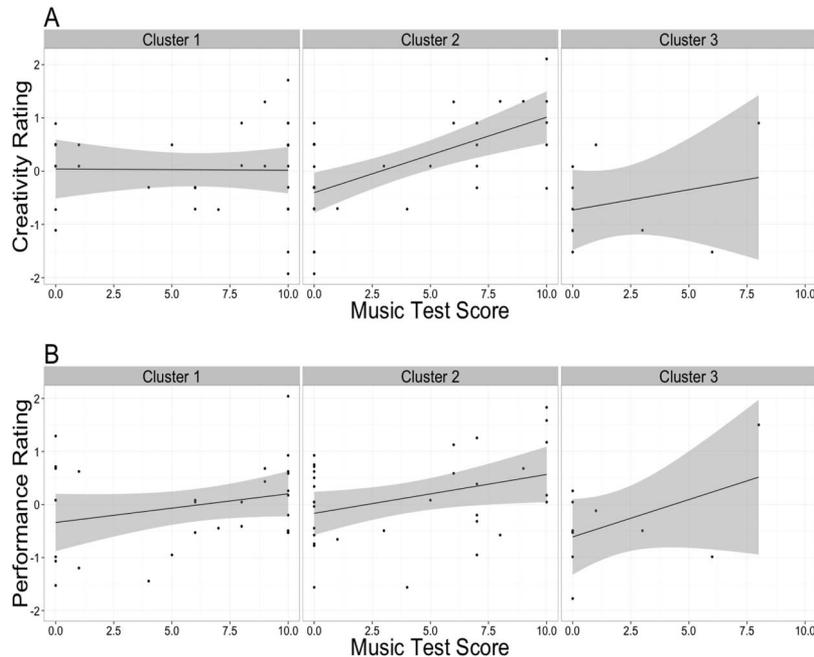


Figure 4. Data from two univariate linear models examining the influence of music test scores, cluster membership, and the interaction between music test scores and cluster membership on ratings of creativity (A) and performance (B). Panel A shows a significant interaction between music test scores and cluster membership when predicting creativity. That is, music test scores predict creativity, but only for cluster 2. Panel B shows a significant main effect of music test scores only. The y-axes for Panel A and Panel B represent scaled, centered, and averaged ratings that were originally on a Likert scale from 1–7. The smoothed curves represent the 95% confidence intervals.

an attempt to uncover discrete clusters of improvisations and to characterize each cluster as a stylistic type based on its unique structural musical features. This represents not only a new approach to analyzing improvisation, but an approach for analyzing creative products in general.

Our three-cluster solution was able to effect a significant data reduction of the original 72-improvisation sample, as evidenced by the fact that it accounted for a large proportion of the variance in the dataset (roughly 70%). With regard to structural features, C1 and C2 differed from C3 in that the majority of improvisations in these two clusters used standard, predictable metric patterns, whereas improvisations in C3 used irregular metric structures. Second, C2 differed from C1 and C3 in terms of phrase structure, in that it contained a much higher proportion of repetition, repetition with variation, and greater phrase symmetry than C1 or C3. While the improvisations varied in their tonal properties as well (e.g., major, minor, chromatic), such properties did not contribute strongly to the clustering of the improvisations. Instead, rhythm (metric structure) and phrase structure were the two primary musical features that differentiated the improvisations at the cluster level.

Additional analyses based on expert ratings of the creative quality and performance quality of the improvisations offered further insights into the clusters. Although there were no significant differences between the clusters on expert ratings of creativity or performance quality, higher levels of musical training were found to predict higher performance-quality ratings, regardless of the cluster membership

(Figure 4B). On the other hand, music test scores were only a significant predictor of creativity ratings for improvisations in C2, and not for C1 and C3 (Figure 4A). This suggests that, within C2, the level of perceived creativity of an improvisation increased as the improviser's music test score increased.

Insights for Cognitive Mechanisms

Although the strength of our analysis lies primarily in the descriptive methodology of the improvisation style-types uncovered by the cluster analysis, we can offer some speculations regarding the cognitive mechanisms underlying the process of improvisation, as based on results from the linear models. For example, most improvisations in C2 seemed to demonstrate the processes of “sketch planning” and “evaluative monitoring” (Norgaard, 2011), whereas most improvisations in C1, and especially those in C3, seemed not to. Because the improvisations in C2 had highly symmetrical phrases, in addition to repeated phrases that contained structural variations of the phrases that preceded them, the improvisers seemed to demonstrate that they were able to plan ahead (sketch planning) and remember what they had already improvised so as to repeat and modify these ideas if they found them to be successful (evaluative monitoring). These are strategic skills used by expert improvisers (Norgaard, 2011), and so it might have been the case that participants in C2 were more experienced improvisers than participants in C1 or C3. Expertise at musical improvisation may lead to automated sensorimotor (e.g., muscle memory) processes for improvising (Fidlon, 2011). This could allow

the more experienced improvisers to use their limited working memory resources to focus on the overall improvisational structure, allowing their implicit skills (e.g., prelearned motor programming) to manage the note-to-note choices (Berkowitz, 2010). Additionally, individuals in C2 may have had stronger skills for executive function (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001), which could have allowed them to determine the general direction of the improvisation early during their performance and thereby focus on note-to-note changes.

On the other hand, more-novice improvisers may, by necessity, devote much of their executive function and working memory to these note-level, rather than phrase-level, decisions because of a lack of experience with performing and planning ahead during improvising. Improvisations in C1 and C3 seemed to demonstrate these characteristics, because they contained more phrase-to-phrase diversity (i.e., little or no repetition). Such increased diversity may be representative of style choices alone, but could also suggest that the participants may have focused on note-to-note choices, rather than overall structural planning.

Results from the rating analyses showed that there were no significant differences between the improvisation clusters with regard to performance ratings or creativity ratings. However, there was a significant interaction such that music test scores predicted creativity ratings for improvisations in C2. Because improvisers in C2 tended to use musical structures that suggested a potential proficiency in processes like sketch planning and evaluative monitoring, it is unsurprising that there would be a positive relationship with creativity in this cluster, in contrast to C1 and C3. This interaction might suggest that participants with higher test scores were better able to structure their improvisations than people with lower scores. Another interpretation, one related to the repetition of phrases, would argue that those participants with higher test scores may have used repetition and variation in more sophisticated ways than those with lower music test scores.

Musical Creativity in Novices

The field of improvisation has been strongly dominated by the analysis of experts, and so the current work represents one of the first analyses of musical improvisation in adult novices. Whereas Mace and Ward's (2002) seminal study of professional visual artists recommended using professional subjects for the study of creativity, because of factors related to commitment, motivation, and effort, it is also important to examine novice creators, especially given the fact that nearly everyone can exhibit creative behavior in some form (i.e., "everyday" creativity; Richards, 2007; Runco, & Bahleda, 1986). Creative improvisation occurs in everyday activities, such as conversation (Sawyer, 1999, 2000), and the melodic improvisations in our sample may be representative of the spontaneous singing that occurs in daily life (e.g., singing in the shower or in the car). Singing is a universal human activity that is considered to be one of the most natural means of human expression (Lomax, 1968; Nettl, 2015). For example, children have a natural tendency to sing spontaneously early in development (Dalla Bella, Giguère, & Peretz, 2007), and several studies have sought to analyze the products of children's vocal improvisations (e.g., Campbell, 1998; Cohen, 2011; Moog, 1976; Moorhead & Pond, 1978; Raju & Ross, 2012; Raju et al., 2015; Sundin, 1997; Young, 2002). Even though singing begins naturally as a process of vocal play during the early years of life, not all individuals

continue to sing or make music after childhood (Nordoff & Robbins, 1983). When examining everyday creativity, it would be useful to apply the current approach to children's improvisations. In sum, because of the lack of research on vocal improvisation, we feel that melodic improvisation by novice improvisers should be a topic for further analysis by researchers in the field of creativity.

Limitations

Choice of classification features. CantoCore (Savage et al., 2012) was modeled after Lomax's Cantometrics classification scheme (Lomax, 1968, 1976). However, unlike Cantometrics, it is restricted to the *structural* features of songs, rather than features associated with performance style (Savage et al., 2012). This makes CantoCore a reasonable tool for studying vocal musical improvisations, because the study of musical structure is the dominant focus when analyzing improvisations. However, future research would benefit from the development and validation of a classification scheme specifically tailored to the structural analysis of musical improvisation. Such a scheme could be informed by research that analyzes the dimensions of musical improvisation (Biasutti & Frezza, 2009), concepts about what makes musical improvisation creative (Jordanous & Keller, 2012), factors associated with improvisational expertise (Wopereis et al., 2013), and factors that influence improvisation achievement (Madura Ward-Steinman, 2008).

Discreteness of clusters. The method of Hwang et al. (2006) demonstrated promise as an approach for the stylistic analysis of musical improvisation, given the existence of a valid classification scheme of musical features. That said, an important limitation in the interpretation of clusters when using this approach with a relatively homogeneous dataset like ours is that there are no clear criteria for making decisions about which musical features are associated most strongly with each cluster, except for visual inspection of the MCA plot. In cases where there are many shared features between samples, it may be challenging to characterize cluster-specific styles using this approach. In the case of expert improvisers, for example, it may be reasonable to assume that a large majority of musical features within a corpus may be shared between individuals' improvisations, which could make it challenging to isolate style-types. However, this simply suggests that experiments should be guided by relevant research questions.

For example, jazz pianists produce different types of improvisations when they improvise in familiar versus unfamiliar keys (Goldman, 2013). The analytical approach we used is ideal for experiments about improvisation that include experimental manipulations, where improvisation can occur in different conditions. Therefore, the current paper has demonstrated that, even with relatively homogenous data, it is possible to uncover meaningful musical features that differentiate subgroups within the data.

Rater disagreement. Although interrater reliability for the CantoCore codings was similar to that reported in previous studies, it was lower than we had anticipated. This suggests that additional training with CantoCore may be necessary to improve interrater reliability. Another possibility is to have expert improvisers perform the codings, rather than researchers with music training. Another limitation was that we had to discard one of the expert rater's data for the creativity-rating analysis. The literature on the CAT suggests that high levels of interrater reliability should be obtained using this method (Baer & McKool, 2009). Given the

evidence that rater 3 was not in agreement with raters 1 and 2 for the creativity ratings, and based on the fact that rater 3 had a dissimilar musical background from raters 1 and 2, we felt justified in discarding data from rater 3.

Identifying mechanisms. The strength of our analysis lies in the classification-and-cluster-approach to identifying and describing style-types in a corpus of improvisations. Because our improvisation task was open-ended, it did not permit us to draw any strong conclusions about mechanisms. Future work could combine our classification-and-clustering approach with experimental manipulations that attempt to influence the cognitive mechanisms of improvisation, including such factors as working memory (Baddeley, 2003), divergent thinking (Baer, 1996), and executive function (Miyake et al., 2001). For example, Beaty et al. (2013) showed that divergent-thinking scores for a sample of jazz performance students significantly predicted the creativity ratings of their improvisations, but this outcome was not discussed in relation to the creative products (i.e., the musical features generated). The combination of experimental manipulation with our methodology would permit the joint understanding of cognitive mechanisms with the descriptive analysis of the products of improvisation, including structural features. Finally, our music test score was only a very basic measure of musical ability, and future work should consider implementing measures of musical training that are more fine-grained (Müllensiefen, Gingras, Musil, & Stewart, 2014; Ollen, 2006) and that take into account the quantity/quality of improvisation experience.

Conclusions

The results of the current analysis provide support for the classification-and-cluster approach to analyzing musical improvisation. This analytical framework shows promise for the analysis of creative products not just in musical improvisation but across many domains of creativity, including the products of both spontaneous and long-term creativity. The method can be readily applied to other corpus analyses, such as those of jazz instrumental improvisations (e.g., Norgaard, 2014) or children's vocal improvisations (Raju & Ross, 2012). In addition, our classification/clustering method can be combined with experimental manipulations of the creative task (e.g., Fidlou, 2011; Goldman, 2013), individual measurements of cognitive factors (e.g., Beaty et al., 2013), and/or ratings of creativity. Such approaches permit more-direct inferences of the mechanisms of improvising than is possible without such manipulations or measures. However, even in the absence of such experimental manipulations or cognitive factors, a post hoc analysis of the clusters, based on the structural features of the products, permits reasonable speculations to be made about the generative processes involved in improvising.

References

- Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 651–657). Thousand Oaks, CA: Sage.
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*, 997–1013. <http://dx.doi.org/10.1037/0022-3514.43.5.997>
- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839. <http://dx.doi.org/10.1038/nrn1201>
- Baer, J. (1996). The effects of task-specific divergent thinking training. *The Journal of Creative Behavior*, *30*, 183–187. <http://dx.doi.org/10.1002/j.2162-6057.1996.tb00767.x>
- Baer, J., & McKool, S. S. (2009). Assessing creativity using the consensual assessment. In C. S. Schreiner (Ed.), *Handbook of research on assessment technologies, methods, and applications in higher education* (pp. 65–77). Hershey, PA: IGI Global. <http://dx.doi.org/10.4018/978-1-60566-667-9.ch004>
- Banerjee, M., Capozzoli, M., McSweeney, L., & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics*, *27*, 3–23. <http://dx.doi.org/10.2307/3315487>
- Beaty, R. E. (2015). The neuroscience of musical improvisation. *Neuroscience and Biobehavioral Reviews*, *51*, 108–117. <http://dx.doi.org/10.1016/j.neubiorev.2015.01.004>
- Beaty, R. E., Smeekens, B. A., Silvia, P. J., Hodges, D. A., & Kane, M. J. (2013). A first look at the role of domain-general cognitive and creative abilities in jazz improvisation. *Psychomusicology: Music, Mind, and Brain*, *23*, 262–268. <http://dx.doi.org/10.1037/a0034968>
- Benzécri, J. P. (1979). Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire [The calculation of the inertia ration in the analysis of a questionnaire]. *Les Cahiers de l'Analyse des Données*, *4*, 377–378.
- Berkowitz, A. (2010). *The improvising mind: Cognition and creativity in the musical moment*. Oxford, UK: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199590957.001.0001>
- Biasutti, M., & Frezza, L. (2009). Dimensions of music improvisation. *Creativity Research Journal*, *21*, 232–242. <http://dx.doi.org/10.1080/10400410902861240>
- Brown, S., Savage, P. E., Ko, A. M.-S., Stoneking, M., Ko, Y.-C., Loo, J.-H., & Trejaut, J. A. (2014). Correlations in the population structure of music, genes and language. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20132072. <http://dx.doi.org/10.1098/rspb.2013.2072>
- Campbell, P. S. (1998). The musical cultures of children. *Research Studies in Music Education*, *11*, 42–51. <http://dx.doi.org/10.1177/1321103X9801100105>
- Carson, S. (2006). *Creativity and mental illness*. Invitational panel discussion hosted by Yale's Mind Matters Consortium, New Haven, CT, April 19, 2006.
- Cohen, A. J. (2011). Creativity in singing: Universality and sensitive developmental periods? In D. Hargreaves, D. Miell, & R. MacDonald (Eds.), *Musical imaginations: Multidisciplinary perspectives on creativity, performance and perception* (pp. 173–189). New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199568086.003.0011>
- Cohen, A. J. (2015). The AIRS Test Battery of Singing Skills: Rationale, item types, and lifespan scope. *Musicae Scientiae*, *19*, 238–264. <http://dx.doi.org/10.1177/1029864915599599>
- Cohen, A. J., Armstrong, V. L., Lannan, M. S., & Coady, J. D. (2009). A protocol for cross-cultural research on the acquisition of singing. *Annals of the New York Academy of Sciences*, *1169*, 112–115. <http://dx.doi.org/10.1111/j.1749-6632.2009.04771.x>
- Collins, D. (2007). Real-time tracking of the creative music composition process. *Digital Creativity*, *18*, 239–256. <http://dx.doi.org/10.1080/14626260701743234>
- Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *The Journal of the Acoustical Society of America*, *121*, 1182–1189. <http://dx.doi.org/10.1121/1.2427111>
- De Dreu, C. K., Nijstad, B. A., Baas, M., Wolsink, I., & Roskes, M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin*, *38*, 656–669. <http://dx.doi.org/10.1177/0146167211435795>
- Everitt, B. S., & Horton, T. (2014). *A handbook of statistical analyses using R* (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.

- Fidlon, J. D. (2011). *Cognitive dimensions of instrumental jazz improvisation*. (Doctoral dissertation). University of Texas Thesis and Dissertation Repository.
- Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research, and applications*. Cambridge, MA: MIT Press.
- Goldman, A. (2013). Towards a cognitive-scientific research program for improvisation: Theory and an experiment. *Psychomusicology: Music, Mind, and Brain*, 23, 210–221. <http://dx.doi.org/10.1037/pmu0000020>
- Greenacre, M. (2007). *Correspondence analysis in practice* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC. <http://dx.doi.org/10.1201/9781420011234>
- Hartigan, J. A., & Wong, M. A. (1979). A k-means clustering algorithm. *Journal of the Royal Statistical Society Series A*, 28, 100–108.
- Hickey, M., & Lipscomb, S. D. (2006). How different is good? How different is different? The assessment of children's creating musical thinking. In I. Deliege & G. A. Wiggins (Eds.), *Musical creativity: Multidisciplinary research in theory and practice* (pp. 97–110). New York, NY: Psychology Press.
- Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, 105, 302–308. <http://dx.doi.org/10.1037/0033-2909.105.2.302>
- Husson, F., & Josse, J. (2014). Multiple correspondence analysis. In J. Blasius & M. Greenacre (Eds.), *Visualization and verbalization of data* (pp. 165–184). Boca Raton, FL: CRC Press.
- Hwang, H., Dillon, W. R., & Takane, Y. (2006). An extension of multiple correspondence analysis for identifying heterogeneous subgroups of respondents. *Psychometrika*, 71, 161–171. <http://dx.doi.org/10.1007/s11336-004-1173-x>
- Järvinen, T. (1995). Tonal hierarchies in jazz improvisation. *Music Perception*, 12, 415–437. <http://dx.doi.org/10.2307/40285675>
- Joorabchi, M. E., Mesbah, A., & Kruchten, P. (2013). Real challenges in mobile app development. In *Empirical Software Engineering and Measurement*, 15–24. ACM/IEEE International Symposium.
- Jordanous, A., & Keller, B. (2012). What makes musical improvisation creative? *Journal of Interdisciplinary Music Studies*, 6, 151–175.
- Kaufman, J. C., Baer, J., & Cole, J. C. (2009). Expertise, domains, and the consensual assessment technique. *The Journal of Creative Behavior*, 43, 223–233. <http://dx.doi.org/10.1002/jc.2162-6057.2009.tb01316.x>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. <http://dx.doi.org/10.2307/2529310>
- Lomax, A. (1968). *Folk song style and culture*. Washington, DC: American Association for the Advancement of Science.
- Lomax, A. C. (1976). *Cantometrics: An approach to the anthropology of music*. Berkeley, CA: University of California, Extension Media Center.
- Mace, M.-A., & Ward, T. (2002). Modeling the creative process: A grounded theory analysis of creativity in the domain of art making. *Creativity Research Journal*, 14, 179–192. http://dx.doi.org/10.1207/S15326934CRJ1402_5
- Madura Ward-Steinman, P. (2008). Vocal improvisation and creative thinking by Australian and American university jazz singers: A factor analytic study. *Journal of Research in Music Education*, 56, 5–17. <http://dx.doi.org/10.1177/0022429408322458>
- Meyer, L. B. (1989). *Style and music: Theory, history, and ideology*. Chicago, IL: University of Chicago Press.
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130, 621–640. <http://dx.doi.org/10.1037/0096-3445.130.4.621>
- Moog, H. (1976). *The musical experience of the pre-school child*. London, UK: Schott Music.
- Moorhead, G. E., & Pond, D. (1978). *Music of young children*. Santa Barbara, CA: Pillsbury Foundation for the Advancement of Music Education.
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9, e89642. <http://dx.doi.org/10.1371/journal.pone.0089642>
- Nettl, B. (2015). *The study of ethnomusicology: Thirty-three discussions*. Urbana, IL: University of Illinois Press.
- Nordoff, P., & Robbins, C. (1983). *Music therapy in special education*. St-Louis, MO: MMB Music.
- Norgaard, M. (2011). Descriptions of improvisational thinking by artist-level jazz musicians. *Journal of Research in Music Education*, 59, 109–127. <http://dx.doi.org/10.1177/0022429411405669>
- Norgaard, M. (2014). How jazz musicians improvise: The central role of auditory and motor patterns. *Music Perception*, 31, 271–287. <http://dx.doi.org/10.1525/mp.2014.31.3.271>
- Ollen, J. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* (Doctoral dissertation). OhioLINK Electronic Theses and Dissertation Center.
- Pan, B.-Y., & Cohen, A. (2012). The AIRS on-line test battery of singing skills. In Abstracts of the 2010 CSBBS annual meeting. *Canadian Journal of Experimental Psychology*, 64, 286–329.
- Pan, B.-Y., & Cohen, A. J. (2016). *AIRS-Test: A Web-based system for generating on-line experiments requiring audiovisual presentation and recording*. Manuscript submitted for publication.
- Pressing, J. (1998). Error correction processes in temporal pattern production. *Journal of Mathematical Psychology*, 42, 63–101. <http://dx.doi.org/10.1006/jmps.1997.1194>
- Pressing, L. (1988). Improvisation: Methods and models. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 129–178). Oxford, UK: Clarendon Press.
- R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Raju, M., & Ross, J. (2012). Adaption to Estonian children of the protocol for cross-cultural research in singing. *Trames*, 16, 125–144. <http://dx.doi.org/10.3176/tr.2012.2.02>
- Raju, M., Välja, L., & Ross, J. (2015). Estonian children's improvisational songs, the nature of performance and songs' coherence with the Western tonal musical canon. *Musicae Scientiae*, 19, 282–300. <http://dx.doi.org/10.1177/1029864915598663>
- Richards, R. (2007). Everyday creativity and the arts. *World Futures*, 63, 500–525. <http://dx.doi.org/10.1080/02604020701572707>
- Runco, M. A., & Bahleda, M. D. (1986). Implicit theories of artistic, scientific, and everyday creativity. *The Journal of Creative Behavior*, 20, 93–98. <http://dx.doi.org/10.1002/jc.2162-6057.1986.tb00423.x>
- Rzeszutek, T., Savage, P. E., & Brown, S. (2012). The structure of cross-cultural musical diversity. *Proceedings of the Royal Society B: Biological Sciences*, 279, 1606–1612. <http://dx.doi.org/10.1098/rspb.2011.1750>
- Savage, P. E., & Brown, S. (2014). Mapping music: Cluster analysis of song-type frequencies within and between cultures. *Ethnomusicology*, 58, 133–155. <http://dx.doi.org/10.5406/ethnomusicology.58.1.0133>
- Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 8987–8992. <http://dx.doi.org/10.1073/pnas.1414495112>
- Savage, P. E., Merritt, E., Rzeszutek, T., & Brown, S. (2012). CantoCore: A new cross-cultural song classification scheme. *Analytical Approaches to World Music*, 2, 87–137.

- Sawyer, R. K. (1999). Improvised conversations: Music, collaboration, and development. *Psychology of Music, 27*, 192–205. <http://dx.doi.org/10.1177/0305735699272009>
- Sawyer, R. K. (2000). Improvisation and the creative process: Dewey, Collingwood, and the aesthetics of spontaneity. *The Journal of Aesthetics and Art Criticism, 58*, 149–161. <http://dx.doi.org/10.2307/432094>
- Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., . . . Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts, 2*, 68–85. <http://dx.doi.org/10.1037/1931-3896.2.2.68>
- Sundin, B. (1997). Musical creativity in childhood: A research project in retrospect. *Research Studies in Music Education, 9*, 48–57. <http://dx.doi.org/10.1177/1321103X9700900106>
- Wopereis, I. G. J. H., Stoyanov, S., Kirschner, P. A., & Van Merriënboer, J. J. G. (2013). What makes a good musical improviser? An expert view on improvisational expertise. *Psychomusicology: Music, Mind, and Brain, 23*, 222–235. <http://dx.doi.org/10.1037/pmu0000021>
- Young, S. (2002). Young children's spontaneous vocalizations in free-play: Observations of two-to three-year-olds in a day-care setting. *Bulletin of the Council for Research in Music Education, 152*, 43–53.

Received April 5, 2016

Accepted May 18, 2016 ■