AN ANALYSIS OF STYLE-TYPES IN MUSICAL IMPROVIATION USING CLUSTERING METHODS

AN ANALYSIS OF STYLE TYPES IN MUSICAL IMPROVISATION USING CLUSTERING METHODS

By BLAIR KENNETH ELLIS, B.Mus., B.A.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the

Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2015)
McMaster University
Psychology, Neuroscience, and Behavior

TITLE: An analysis of style types in musical improvisation using clustering methods

AUTHOR: Blair Kenneth Ellis
B.Mus. (Mount Allison University)
B.A. (Mount Allison University)

SUPERVISOR: Dr. Steven Brown

NUMBER OF PAGES: x, 52

Abstract

Research on creativity 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 output of improvisation due to the variable nature of the products. In the case of musical improvisation, structural approaches have used methodologies like musical transcription to look for recurring or variable musical features across a corpus of improvisations, while creativity-centered approaches have had experts make ratings of the novelty of the improvisations. One important concept missing from many analyses 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. In order to look at the potential of such methods, we examined a corpus of 72 vocal melodic improvisations produced by novice improvisers. We first classified the melodies acoustically using a multidimensional musical-classification scheme called CantoCore, which coded the melodies for 19 distinct features of musical structure. We next employed the simultaneous use of multiple correspondence analysis (MCA) and k-means cluster analysis with the data, and obtained three relatively discrete clusters of improvisations. Stylistic analysis of these clusters revealed that they differed in key 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.

Acknowledgements

My time as a graduate student at McMaster has been an incredible learning experience. Before arriving in Hamilton, I could not have imagined learning so much in only two short years, and it would have been impossible without the support of friends, family, and colleagues. PNB is filled with remarkable people and brilliant ideas, and I'm proud to have been a part of this family for a short time.

I would like to thank my parents, Kenneth and Anne-Marie, and my sisters Amy Lynn and Cassie for their unconditional love and support during my university education. I'm sure it has been challenging for them to follow me during what must seem like a strange creative path, but I have never questioned their support. Also, the love I have felt from my partner Jocelyn and the whole Wichtel family during my time in graduate school will always be remembered.

I owe a debt of gratitude to Annabel Cohen for guiding me through my introduction to the world of music psychology research from my undergraduate years through to today, and for access to the data used in this thesis. Thank you to my supervisor, Steven Brown, for giving me this opportunity to explore my research interests, sharing his "big-picture" thinking, late night e-mail exchanges, and exciting discussions about our interests in music and artistic creativity.

Thank you to my fellow lab-mates. Thank you Michel Belyk for dealing with my endless questions; I assure you that I've gotten better at only asking those questions that matter most. Thank you to Léa Chauvigné for your friendship and for your help with MATLAB, and to Ye Yuan and Matthew Berry for filling the lab with an air of positivity.

This thesis would not have been possible without the gracious support of Patrick Savage and Heungsun Hwang. Thank you for being a great teacher, Pat. You taught me so much about music classification and multivariate data analysis, and you made yourself available upon request despite your busy family life, the birth of your daughter, and your ambitious pursuits in the field of comparative musicology. Thank you, Heungsun, for agreeing to collaborate with me, and for your patience and e-mail correspondence as I learned how to apply your method with my data. Our discussions at McGill inspired me to develop a deeper understanding of statistics, and I will never forget your generosity and kindness as you guided me in this pursuit.

Bing-Yi Pan and Ross Dwyer – you both showed me a great deal of support during my time at McMaster. Whether it was advice over phone calls, or running around at UPEI to lend a helping hand while I was in Hamilton, I can't thank you enough. Kate Thompson, thank you for the time you devoted to learning how to code the data as an inter-rater, and I wish you luck in your own graduate pursuits.

My deep appreciation also goes to all of my peers in PNB. Whether it was our regular afternoon coffee breaks across Main Street, pizza nights over our favourite beverage, de-stress sessions over a few pints at the Judge, or during our many stimulating and/or ridiculous discussions, I know I've made lifelong friends here at McMaster.

Table of Contents

Preface

The following thesis contains three chapters. Chapter 1 is an introduction that describes relevant literature and the motivation to conduct the study outlined in this thesis. Chapter 2 reports the methodology and results of two different analyses: first, a classification and cluster analysis of solo vocal improvisation data using multivariate statistical techniques; and second, an analysis of the same solo vocal improvisation data in reference to the improvisers' music-theory knowledge and to expert raters' perceptions of the performance quality and creativity of the samples. Chapter 3 presents a general discussion of the findings of the current study. This thesis will be condensed and submitted for publication.

List of Figures

CHAPTER 2

APPENDIX

List of Tables

CHAPTER 2

APPENDIX

CHAPTER 1: Introduction

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 creative generation of novel products can be examined along either the long times scales required to produce apps (Joorabchi, Mesbah, & Kruchten, 2013), musical compositions (Collins, 2007), and scientific theories (Finke, Ward, & Smith, 1992; Mace & Ward, 2002) or 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 processes of spontaneous creativity (Beaty, Smeekens, Silvia, Hodges, & Kane, 2013; Bengtsson, Csíkszentmihályi, & Ullén, 2007; Berkowitz & Ansari, 2008, 2010; Fidlon, 2011; Goldman, 2013; Järvinen & Toiviainen, 2000; Järvinen, 1995; Limb & Braun, 2008; Lopata, 2014; Norgaard, 2011, 2014; Norgaard, Spencer, & Montiel, 2013; Pfleiderer & Frieler, 2010). Unlike work on divergent thinking (Silvia et al., 2008) the research literature on improvisation has typically looked at specialists with a lifetime of training in performance and a professional training in improvisation itself, most notably jazz instrumentalists and singers. 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 pre-learned musical structures at stylistically appropriate times, and use perceptual feedback and error correction to monitor and adjust differences between intention and actualization. The improviser may also choose to incorporate material that is completely unrelated to the present context (Norgaard, 2014), further diversifying the product.

Because of the diverse nature of these products, the analysis of improvisational performance has provided many analytical challenges for researchers and theoreticians. Previous research in this area has used music-analytic approaches to examine the stylistic features of improvisations in order to gain insight into the underlying cognitive processes during improvisation (Järvinen & Toiviainen, 2000; Järvinen, 1995; Norgaard, 2011; Norgaard, 2014; Norgaard et al., 2013; Pfleiderer & Frieler, 2010; Raju & Ross, 2012), and some studies have used a cognitive experimental approach (Fidlon, 2011; Goldman, 2013). Studies like these have attempted to quantify the structure of music improvisations in different ways, but none of them have captured a systematic multidimensional musical signature of each sample being analyzed. For example, the 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), the statistical distribution of_tone choices across metrically salient beats (Järvinen, 1995), the emphasis of tone choices across other metrical structures aside from metrically salient beats (Järvinen & Toiviainen, 2000). In addition, these studies, as well as more extensive multivariate analyses examining the products of improvisation (Pfleiderer & Frieler, 2010), have relied heavily on musical transcriptions or MIDI recordings. However, few attempts have been made to analyze the products of musical improvisation in a more holistic multivariate manner that does not require transcription (Hickey & Lipscomb, 2006; Madura Ward-Steinman, 2008; Raju & Ross, 2012).

Our approach attempts to demonstrate a method that can be used to understand the structure of a diverse corpus of improvisations. It takes advantage of methods coming from classification and cluster analyses. In particular, we are proposing to classify a sample of musical improvisations using the musical classification scheme "CantoCore" that was developed by Savage, Merritt, Rzeszutek, & Brown (2012). This is a multidimensional coding scheme that contains 26 structural characters of musical structure that span the broad domains of rhythm, pitch, syllable, texture, and form. Each of the 26 characters contain between three and six character-states. This scheme was previously used – in combination with multidimensional scaling and k-means cluster analysis – 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). Since the coding scheme was multidimensional, the observed clusters differed from one another in a multidimensional manner. In other words, the clusters represented 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 stylistic song-types derived from classification and cluster analyses.

The goal of the current study was to apply this type of stylistic analysis to a corpus of musical improvisations. This follows in the tradition of Meyer's (1989) classic treatise on musical style, which defined style as "…a replication of patterning, whether in human behavior or in the artifacts produced by human behavior, that results from a series of choices made within some set of constraints" (p. 1). However, we decided to use a more sophisticated clustering procedure for the current study. In particular, we took advantage of the power of the statistical technique known as multiple correspondence analysis (MCA; Abdi & Valentin, 2007). MCA stems from the field of geometric data analysis that began in the 1960s with the work of J.-P. Benzécri (for brief historical overview, see LeRoux & Rouanet, 2010). It is a powerful data-reduction technique that can be used to describe the underlying structure of multivariate categorical data. In its current implementation, it can be thought of as a principal component analysis (PCA), but one that is used for categorical data, rather than continuous data (Abdi & Valentin, 2007). Despite its strengths, MCA is rarely discussed and is very underused (LeRoux & Rouanet, 2010). In fact, the present study is the first attempt at analyzing music using MCA, which we consider to be an informative approach for the classification and stylistic analysis of music. The specific statistical framework that we adopted in this study is outlined in Hwang, Dillon, and Takane, (2006), and involves the simultaneous implementation of MCA and k-means cluster analysis (Hartigan & Wong, 1979). K-means cluster analysis is a 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 an optimal number of relatively discrete style-clusters. The final result of this method allows for an integrated graphical display (i.e., a 2D or 3D plot) that permits easy interpretation of relatively discrete cluster-based improvisation styles within the sample, as well as the interrelationships among these styles based on the musical features on which they were coded (Hwang et al., 2006).

The principal objective of the current project was to apply a combination of classification and clustering methods to a corpus of improvisations to see if the data-reduction potential of these methods could shed light on the structural organization of the corpus and provide insight into the generative mechanisms of improvising. In other words, we wanted to examine whether we could identify subsets of stylistic types (cantogroups) within the corpus as well as provide insights into the different manners of improvising that underlie these different stylistic types. For the corpus, we used a set of 72 vocal melodic improvisations generated as part of the AIRS (Advancing Interdisciplinary Research in Singing) Test Battery of Singing Skills (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 in order to generate clusters of improvisational style. The goal then was to use the multivariate plot that results from the analysis in order 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 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 samples for their level of creativity and performance quality in order to look for correlations between the structural features of the clusters and external assessments of creativity.

CHAPTER 2: Methods, Data Analysis, and Results

Methods

The Improvisation Data

The vocal improvisation data were collected using the online interactive AIRS Test System (Pan & Cohen, 2010). This system is an automated, audio-visual, computer-based version of the original in-person interview version of the AIRS Test Battery of Singing Skills (Cohen et al., 2009), which assesses 11 main components of human singing skills. Although all participants in this study completed the 11-component test, the current study only analyzed component No. 9, namely the vocal improvisation task (described in detail below), which consisted of each participant creating two melodies, one with words and one without. Participant data from the test battery was uploaded to an online database, which was accessible by researchers affiliated with the AIRS project for research purposes.

Participants. At the time that the study began, 110 adult participants (67 females, 43 males; age: M=31.70, SD = 17.29) living in Charlottetown, PE or the surrounding area had participated in the AIRS Test Battery. They were recruited through posters or word of mouth. After the experiment was done, all participants completed an online e-mail survey created at UPEI about their language and musical background, including a 10-point test assessing knowledge of music theory and music reading skills (see Table S1 in the Appendix). Of the 110 participants, data from 38 participants were excluded. Reasons for exclusion included: instances where participants generated improvisations that copied pre-existing music presented earlier in the test battery or well-known songs that existed in pop culture; 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 recorded 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) 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; musicianship: 33 musicians, 39 non-musicians [based on self ratings]; age: M=30.79, SD=16.68). 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 AIRS Test Battery of Singing Skills database. Participants were guided to the sound-attenuated chamber and were invited to relax and to acclimate to the space before reading and signing a consent form. They were also asked to fill out a form to collect personal and contact information. This took about 10 minutes, during which time the participants did not listen to any music. The duration of the entire 11-component test battery was approximately 30 minutes. An online survey of language and music background, including a 10-point test assessing basic knowledge of music theory (Table S1 in the Appendix), was sent to each participant by e-mail after the experiment.

The improvisation task. Data from component No. 9 of the AIRS Test Battery of Singing Skills were analyzed for the current study. During this task, participants sat in front of a computer monitor, were presented with a set of four images on the screen (heart, flower, sun, and apple), and were asked to select one. After doing so, the selected image was presented on the screen alone while the participant was asked to mentally create a song inspired by the picture. The images were used to ensure that the participants did not draw a blank when asked to create a spontaneous vocal improvisation, rather than for any theoretical purpose. 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 (i.e., 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 due to our general interest in musical creativity, rather than language creativity. Hence, the final dataset consisted of the 72 improvisations without words.

The Rating Data

In addition to doing a structural analysis of the improvisations produced by our participants, we had professional musicians rate the samples for their level of creativity and performance quality.

Participants. Six university professors (4 male, 2 female; age: M = 56, SD = 11) in 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 pseudo-randomly 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 (BE) 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, a set of MA-10ABK Edirol speakers (9'' x 6'' x 7''), and the presentation software PsychoPy were used to present participants with text-based instructions that allowed them to navigate through the PsychoPy program.

First, the program presented 10 randomly-selected audio recordings of improvisations from the AIRS Test Battery of Singing Skills that were not used in the rating experiment. These melodies were only presented for listening purposes in order to give the raters a sense of what the corpus of improvisations sounded like as a whole. This was done in order 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 the rater was in. While both sets of improvisations were rated (72 without words and 72 with words), only the ratings for the samples without words are reported in the current paper 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 tied to any particular theory of creativity (Baer & McKool, 2009). Using this approach, 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 *musical* 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. In the performance-rating condition, raters were instructed to do precisely the opposite and focus on performance features, but not musical features or creativity. Tasks that require analytical listening, such as having a person focus on particular musical features while ignoring performance features, are common tasks for expert musicians holding advanced degrees. These analytical listening tasks are thus well-practiced skills for such raters.

Data Analysis

Classifying the Improvisation Data

The improvisation audio files. The improvisation samples were downloaded from the online test-battery database and were saved as audio-video .flv files. The files were converted into audio-only .wav files using a script in the Apple terminal. All files were then processed using the application Levelator (version 2.1.1), which adjusted the acoustic amplitude of the audio files. The processing included compression, normalization, and limiting, and adjusted the audio files for amplitude variations from one singer to the next. The musical structure of the improvisations was not influenced by this process, nor were the perceived levels of performance competence for each improvisation. The goal was to decrease the level of feedback or noise created during the recording process so as to ensure that the samples could be played sequentially without requiring repeated manual adjustment of the sound level.

Acoustic classification of the improvisations. In order to classify the improvisations by ear, we used a modified version of the musical-classification scheme CantoCore, which contains 26 characters of musical structure, as organized into the five basic dimensions of rhythm, pitch, syllable, form, and texture (Savage et al., 2012). The modified scheme contained 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 (e.g., item 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 plots in the Results section. These characters of musical structure contain both ordinal and nominal characters-states, as musical characters can be measured either ordinally on a scale from small to large (e.g., interval size or number of pitch classes) or nominally on a presence/absence scale (e.g., does or does not contain an arched melodic contour). When the 19 musical characters are expanded into their respective levels, or coding options, the result is 68 possible character-states (see Table 1 for the character-states used by the coder, and Table S2 in the Appendix for the complete coding scheme including the definitions of all character-states). We will refer to the 68 character-states as "musical features". Items that were not codable or that were arbitrary were marked as "NA" for not applicable and were treated as missing data. The coder (BE) listened to each of the melodies and coded them according to the scheme in Table 1 and Table S2. As is recommended by Savage et al. (2012), the coder listened to each musical sample as often as necessary to arrive at an accurate coding.

Table 1. The modified coding scheme used to code the vocal improvisations. The coder listened to the samples and assigned character-states for each of the 19 characters of the modified CantoCore scheme (see Table S2 in the Appendix 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 purpose of MCA). Graphical labels are presented for the purposes of interpreting the musical character-states in Figure 3.

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	051, 052, 053
6	motivic redundancy	rhythm	i, ii, iii	061, 062, 063
7	durational variability	rhythm	i, ii, iii	071, 072, 073
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

The modified classification scheme. All characters dealing with musical texture (i.e., multiple musical lines) were omitted, since all of the improvisations were solo vocal melodies. The "melisma" character was omitted because all samples were sung without words. They were sung using various syllables, where melismas often sounded arbitrary or were difficult to detect. A new character called "total number of phrases" (line 15) was added to the scheme. Rather than measuring a sample's duration in seconds or minutes, we decided to count the total number of musical phrases in the sample in order to quantify how long each improvisation was. 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 = < 24%, ii = 24%-49%, iii = > 50%). The rationale for including this character in the scheme was that an improvisation that included repetition with variation introduced an additional 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), since it is not possible to have varied repetition without having repetition itself.

Table 2 attempts to explain the logic of including repetition with variation in our coding scheme by presenting hypothetical phrases that illustrate the relationship between characters 16 (phrase repetition) and 17 (repetition with variation). In Table 2, under the column heading "phrase combinations", are hypothetical examples of possible phrase combinations, where individual letters indicate unique musical phrases, and letters with a superscript indicate phrases that contain musical variations of the material that appeared in a previous, similar phrase. The "+" symbols indicate the proportion of between-phrase-type repetition (column 2), between-phrase-type diversity (column 3), or within-phrase-type variation (column 4) occurring in each combination of phrases (where + = low, ++ = moderate, +++ = high, ++++ = very high). For example, for the sample in the first row containing four unique phrases (ABCD), there is no between-phrase-type repetition and therefore no within-phrase-type variation, which also necessarily implies that there is a substantial amount of between-phrase-type diversity. For the sample in the last row (AAAA'), there is a very high degree of phrase-level repetition, no between-phrase-type diversity, and only a small proportion of within-phrase-type variation. "Phrase repetition" (character 16) captures information about both between-phrase-type repetition and between-phrase-type diversity. If an item is coded as having a high degree of phrase repetition, then it will necessarily have a lower degree of phrase diversity, and vice versa. "Repetition with variation" (new character 17), on the other hand, captures information about *within*-phrase-type variation only.

Table 2. The addition of a new character of "repetition with variation" to CantoCore for the current project. In addition to the existing character "phrase repetition" (line 16), "repetition with variation" (line 17) was added. This table presents hypothetical samples and their codings. See text for details.

Phrase Combinations	Between-phrase-type repetition	Between-phrase-type diversity	Within-phrase-type variation
1. ABCD	—	++++	—
2. ABAB	++	++	—
3. ABA^1B^1	++	++	++
4. ABB^1B^1	+++	++	+
5. AAAA	++++	—	—
6. $AAAA^1$	++++	—	+

Inter-rater reliability. In order to measure inter-rater reliability, we had a second rater code 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. Inter-rater reliability was calculated for each individual character of the modified CantoCore coding scheme in two ways. The simplest 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 32.3%. Both values are similar to those observed in previously-published studies using CantoCore (Brown et al., 2014; Savage et al., 2012). According to Landis and Koch (1977), Cohen's Kappa values of 0.21-0.40 are considered "fair" agreement. It is important to note that Cohen's Kappa may not be the best measure of inter-rater agreement (Banjerjee, Capozzoli, McSweeny, & Sinha, 1999).

Clustering the Improvisation Data

Overview. 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. The aim was to describe – in terms of structural musical features – the style of each of the observed improvisation clusters. This approach allows us not only to cluster the samples into relatively discrete style-types but to identify the set of musical features that characterize each style-type, doing so across all 19 characters of musical structure that make up the classification scheme.

The clustering procedure was comprised of three principal steps: 1) MCA was conducted for the purpose of dimensional reduction and to determine the number of dimensions; 2) with a fixed number of dimensions chosen, we determined the optimal number of improvisation-clusters by selecting the number that was the most parsimonious balance between the variance within each cluster and the overall number of clusters; and 3) MCA and k-means cluster analysis were applied simultaneously in a single framework, as described in (Hwang et al., 2006) in order 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 clusters of improvisations that are relatively homogenous within this low-dimensional feature space (Hwang et al., 2006). Table 3 (steps 2-4) presents a sequential breakdown of the overall analysis.

Table 3. A summary of the analysis method for dimensional reduction and cluster analysis.

Step	Order	Level of Analysis	Method	Description
1. Classification of improvisations	1	Pre-data analysis, considering songs independently	CantoCore (Savage et al., 2012)	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 (Greenacre, 2007)	Select the optimal number of dimensions based on maximizing the variance explained and minimizing the number of dimensions (Figure S1 in the Appendix)
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 (Hwang et al., 2008) iteratively with increasing numbers of clusters (1, 2, 3, etc.)	Select the optimal number of improvisation clusters based on minimizing the variance within each cluster and minimizing the number of overall clusters (Figure 2)
4A. Multiple correspondence analysis (MCA)	4A and 4B occur simultaneously	Data analysis; all improvisations in the sample are analyzed together	Data reduction using MCA (Hwang et al., 2008)	Uncover a low-dimensional multivariate musical feature space that explains variance in the data
4B. K-means cluster analysis		Data analysis; all improvisations in the sample are analyzed together	K-means cluster analysis is conducted on improvisation samples with reference to MCA results (Hwang et al., 2006)	Assign improvisations to clusters based on stylistic similarities

Benefits of the simultaneous vs. tandem MCA and cluster analysis. As discussed in Hwang et al. (2006), it is a convenient practice to use a two-step sequential process in classification and cluster analyses (Arabie & Hubert, 1994). The two-step procedure first reduces the data from many dimensions to a low-dimensional representation (e.g., MCA), and then a cluster analysis is conducted to uncover relatively discrete clusters on the basis of the low-dimensional data (e.g., k-means). However, despite the benefits of this approach (Green & Krieger, 1995), it has been highly criticized. For example, the two-step approach does not guarantee that the MCA solutions obtained from step one (dimensional reduction of CantoCore codings via MCA) are optimally chosen for clustering the improvisations in step two (k-means cluster analysis of the dimensionally-reduced CantoCore codings, rather than a cluster analysis of the original codings), because the two techniques are applied independently (Arabie & Hubert, 1994; Chang, 1983; Desarbo, Jedidi, Cool, & Schendel, 1991; De Soete & Carroll, 1994). In other words, in a two-step approach, the MCA solution of the CantoCore codings may explain the variance of the data well, but may not include useful information on clustering individual improvisations. This approach actually has the possibility of masking or distorting cluster structures that exist in the original data (Hwang et al., 2006). On the other hand, the simultaneous approach used in the current analysis (Hwang et al., 2006) aims to obtain both MCA and clustering solutions while considering the techniques simultaneously. The analysis is conducted in such a way as to ensure that the MCA solutions are optimally chosen for clustering, which should result in more-accurate cluster solutions than in the tandem approach. The current study aims to demonstrate the power of this method for the analysis of creative products.

The improvisation data. 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 initially transformed into an "indicator matrix" using MATLAB. In the process of doing this, the 68 character-states of the 19 musical characters (see Table 1) were converted into a binary system. For example, instead of coding character 16 as either "i", "ii", or "iii", it was coded as 0 (absent) or 1 (present) for each of the three levels of "i", "ii", and "iii". This resulted in an indicator matrix of 72 improvisations (rows) by 68 music variable categories (columns) or character-states. This transformation modified the dataset such that all musical features, whether nominal or ordinal, were treated as binary categorical variables for this analysis. For this reason, the potential of multi-coding that CantoCore offers – in which a sample can be coded with more than one character-state for a given musical character – was ignored (Savage et al., 2012). Preprocessing of the data was necessary in order to ensure a stable MCA solution. This consisted of removing all features in which three 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).

Determining the number of dimensions. MCA is a dimensional reduction technique that aims to uncover underlying structures in multivariate categorical data (Greenacre, 2007). MCA was conducted on the final indicator matrix in order to determine the optimal number of 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 with a measurement of variance in MCA – decreased more slowly after the first three values (see Figure S1). Adjusted inertia values are a better approximation of the variance in MCA than are unadjusted inertias (Abdi & Valentin, 2007; Benzécri, 1979). The three adjusted inertias (31.45%, 24.12%, and 13.58%) accounted for 69.15% of the total adjusted inertia (variance) in the data.

Determining the number of clusters. The k-means cluster analysis algorithm, which is used simultaneously with MCA in this analysis to assign improvisations into improvisation-type clusters based on musical features, requires a manually pre-selected number of clusters (Hwang et al., 2006). The optimal number of clusters was selected by examining the scree plot (Figure 1). The number of clusters is selected so as to ensure a parsimonious balance between minimizing the optimization criterion and minimizing the overall number of clusters (Hwang et al., 2006). The optimal number of clusters for this MCA solution was selected as 3 (i.e., c=3, where c refers to the number of clusters), as defined by the elbow (i.e., transition point) of the scree plot (Everitt & Hothorn, 2006).

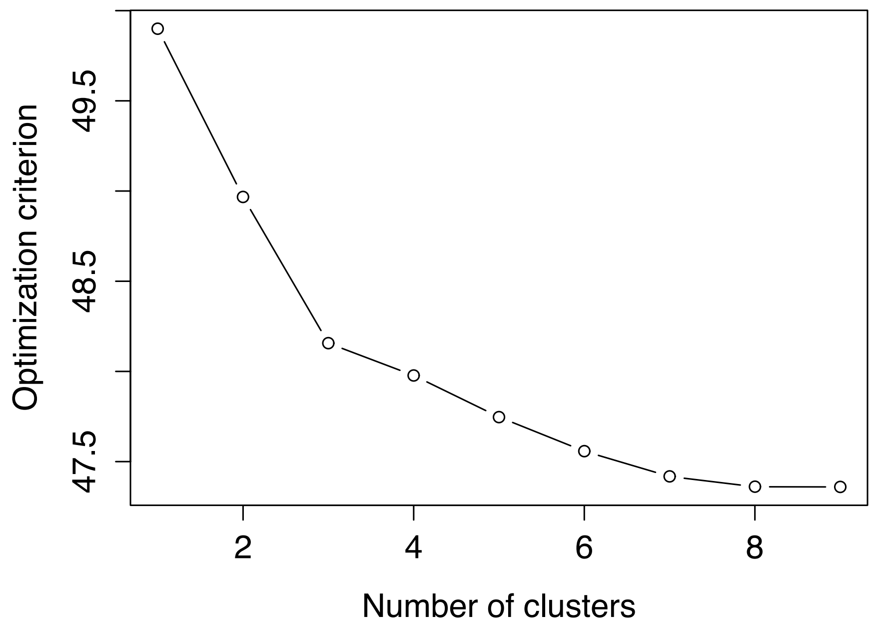

Figure 1. Scree plot for the number of clusters. The number of clusters can be selected by examining the scree plot and ensuring the parsimonious balance between minimizing the optimization criterion and the overall number of clusters. The optimal number of clusters for this MCA solution was 3, as defined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

Simultaneous MCA and k-means cluster analysis. Given the predetermined number of dimensions and number of clusters, 1000 iterations of the simultaneous MCA and k-means cluster analysis was performed with randomly-selected starting locations for the three k-means centroids. Once the iterations were completed, the solution that exhibited the smallest optimization criterion was selected. The result is presented in the form of a three dimensional plot, where the 72 improvisations are clustered into three improvisation styles based on the features that they were coded on using the modified CantoCore scheme. Other non-statistical criteria for determining cluster numbers (e.g., cluster size, interpretability) were also important (Arabie & Hubert, 1994; Wedel & Kamakura, 1998). We determined that a d=3 and c=3 result was an appropriate solution, as based on the scree plots, cluster sizes, and interpretability.

Rating Analyses of Creativity and Performance Quality

Dependent measures. Three musical experts rated each improvisation sample 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 inter-rater reliability between pairs of raters for creativity, and between raters for performance (see Table S3 in the Appendix). 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.

Given the high inter-rater consistency in creativity ratings by 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 performance ratings, except that the single composite rating for performance was based on all three raters, due to their higher inter-rater consistency.

Independent measures. Two different independent measures were used as predictor variables in two one-way ANOVA analyses that aimed to predict ratings of musical creativity and ratings of performance quality separately. First, a music test score out of 10 points was used to assess the level of musicianship for each of the 72 participants (see the Table S1 in the Appendix for the test questions). Second, all participants were assigned a cluster membership (for clusters C1, C2, or C3, as described in the Results section below) based on the clustering procedure discussed in the previous section.

Predicting ratings from cluster membership and music test scores. Two one-way ANOVAs were conducted in order to examine if the scaled, centered, and averaged creativity ratings (Model A) or performance ratings (Model B) could be predicted from cluster membership (C1, C2, C3), music test scores (scored from 0-10), and if there was an interaction between these two factors. Additionally, an independent samples t-test was conducted to test whether or not the difference between the means for music test scores for participants who self-reported themselves as musicians or non-musicians was significantly different from zero.

Results

Simultaneous MCA and K-means Cluster Analysis

Overview. Standard results for MCA analyses consist of two sets of coordinates. Using the terminology of MCA, these are referred to as *individual coordinates* (i.e., the improvisations themselves), based on the rows of the dataset, and *variable coordinates* (i.e., the coded musical features), based on the columns of the dataset (Greenacre, 2007). These coordinates are calculated separately and can then be overlaid in the same representational space on a plot. The interpretation of the results is informed by examining the distance relationship between these two sets of coordinates. When variable coordinates appear close to individual coordinates, it means that both the individual coordinate and the variable coordinate tend to occur together. The distances between individual and variable coordinates are not a direct measure of importance, i.e., where proximity equals importance. Rather, the closeness of variable coordinates to individual coordinates is an emergent result of both sets of variables tending to be highly related (Abdi & Valentin, 2007).

In a special implementation of MCA, we employed a simultaneous k-means cluster analysis for the purpose of uncovering clusters of similar improvisation style within the sample of 72 improvisations. The current analysis will focus on the relationship between the cluster centroid for a given group of improvisations and the variable coordinates (i.e., musical features) that are closest to it. In other words, when a given variable coordinate is closest to the centroid of a particular cluster of improvisations, it is said to be strongly related to that group of improvisations. The results for the individual coordinates (clusters) will be described first, followed by the combination of the individual and variable coordinates (cluster/feature relationships).

Individual coordinates (clusters of improvisations). Figure 2 represents the way that individual improvisations (as shown by circles in the plot) were grouped into three relatively homogeneous clusters of improvisation styles: C1 in red, n=30; C2 in blue,

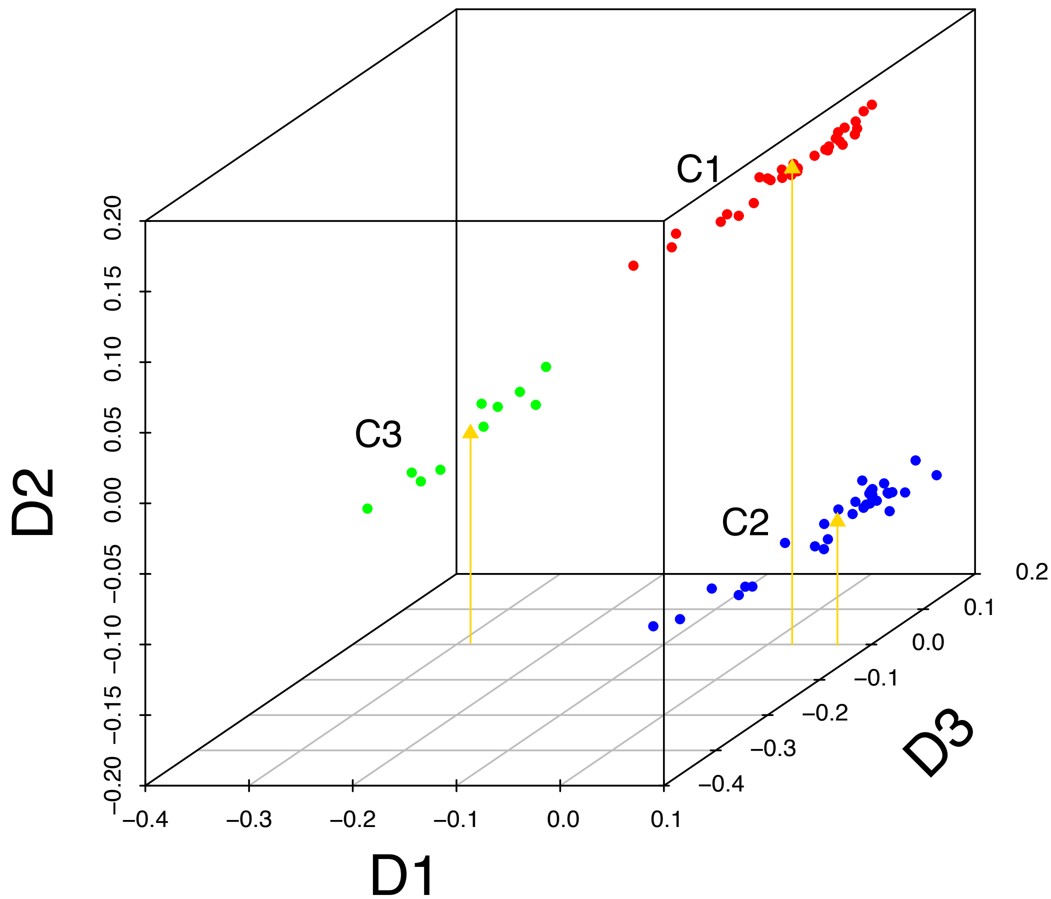

Figure 2. Dimensional plot of the improvisations and the three clusters. The circular points represent individual improvisations, and the three colored clusters (C1, C2, and C3) represent the improvisation styles that were identified in the analysis. The yellow triangles in the center of each cluster represent the cluster centroid for each of the three improvisation styles. It is clear that dimension 3 does not contribute strongly to cluster membership, as the coordinates for each of improvisations are spread out along this dimension. D1, D2, and D3 refer to the first three dimensions of the MCA analysis.

n=32; C3 in green, n=10. Yellow triangles 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 on the variable plot (see the next section entitled "Variable coordinates").

Variable coordinates (musical features). Figure 3 represents the same improvisation clusters as in Figure 2, but in reference to the variable coordinates (open circles; i.e., the 46 character-states of the 19 characters of CantoCore that the improvisations were coded on). 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, while considering any useful information from the other two plots. The interpretation of Figure 3 is descriptive in nature, and the goal is to describe the musical characteristics of the three different clusters of improvisations. As mentioned above, in order to understand the unique musical features of each of the three improvisation clusters, we must consider the distance relationship between the cluster centroids and the musical features (open circles; see Table 1 for labels and Table S2 for definitions of the labels) in Figure 3. The musical features closest to a given cluster centroid (e.g., features "01b", "133" and "171" for cluster C1) tend to be related more strongly to the improvisations in that cluster than to more distant clusters. Although distances between variable coordinates and cluster centroids do not reflect a true measure of proximity or "importance" (Abdi & Valentin, 2007), when a variable coordinate is closer to a given cluster centroid, it does suggest that improvisations in that cluster tend to be coded for that particular feature.

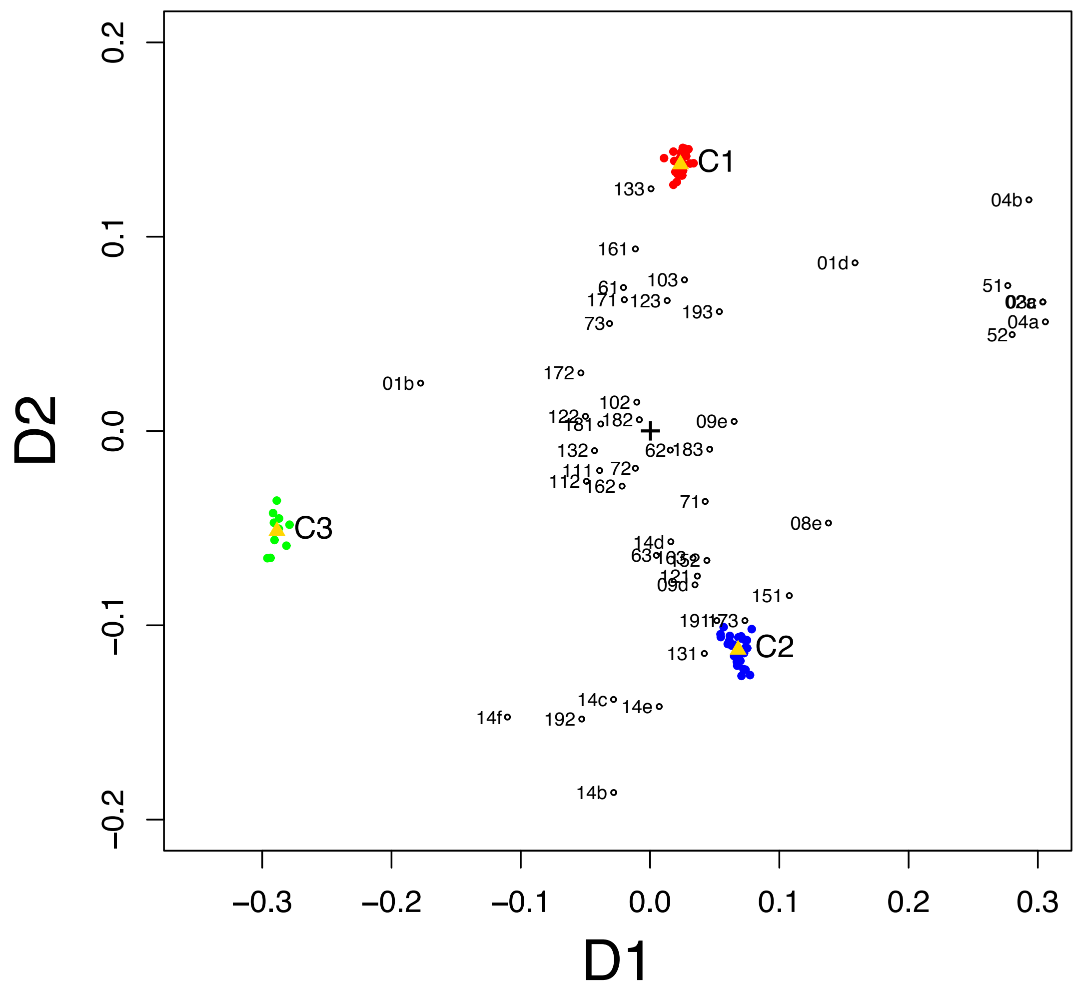

Figure 3. Simultaneous MCA and k-means cluster analysis. The simultaneous MCA/k-means plot has two different types of points: 1) individual coordinates that represent the individual improvisations (colored circles) with their respective cluster centroids (C1, C2, and C3), and 2) variable coordinates (open circles) that represent the character-states, of the 19 CantoCore musical characters (see Table S2 in the Appendix for label definitions). The yellow triangles represent the cluster centroids. 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.

As commonly occurs with cluster analyses, the results in Figure 3 do not demonstrate a non-overlapping solution. In other words, samples in more than one cluster were coded on many of the same musical features. An ideal solution might have been one where there were non-overlapping features between clusters, i.e., where a given cluster might have contained improvisations that were defined by unique musical features that were not present in other clusters (see Figure 1 in Hwang et al., 2006 for a relatively non-overlapping solution). Instead, the current results suggest that the dataset as a whole is rather homogeneous in terms of musical features overall, and that the clusters differ from one another on a small number of musical features.

Interpreting Improvisation “Style” from Cluster Membership

Overview. Due to the overlapping nature of the musical features between clusters, it is challenging to clarify the overall musical structure of each of the three clusters. To examine this, in addition to examining Figure 3 to interpret the data, auxiliary information that resulted form the analysis was also used. Specifically, for each cluster, it is possible to access the percentage of improvisations that were coded as “present” for each of the 46 musical features by using a simple pivot table. It is important to note that this auxiliary information is only meant to supplement the main findings in Figure 3. Figure S2 in the Appendix demonstrates the relative distribution of “presence” for the 46 musical features for each of the three clusters, and Tables 4-6 outline the definitions of each feature.

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" (non-repetitive phrases) than any other features. The improvisations in C1 (n=30) make up 41.7% of the sample. Panel A of Figure S2 represents the relative percentages of presence for each of the 46 musical features for all improvisations in C1, and Table 4 provides a description of those features that were present in at least 50% of the improvisations in C1. Based both on these results and on acoustic analyses of 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 non-repetitive.

Table 4. Musical features of Cluster 1. At least 50% of improvisations contained these features, in descending order of importance (100% → 50% present).

Order	**Label**	**Musical Feature**	**Percentage**
1	**01d**	isometric meter	100%
2	**02a**	number of beats is duple	100%
3	**03c**	iso-divisive sub-beats	100%
4	**08e**	isotonal	97%
5	**51**	little or no syncopation	93%
6	**171**	no repetition with variation	90%
7	**09e**	major iso-modal	83%
8	**161**	non-repetitive phrases	80%
9	**04a**	simple sub-beats	77%
10	**103**	large number of pitch classes	77%
11	**111**	non-hemitonic	73%
12	**122**	medium melodic interval size	70%
13	**132**	moderate pitch ranges	63%
14	**151**	small number of overall phrases	57%
15	**191**	highly symmetrical phrases	57%
16	**72**	moderate degree of durational variability	53%

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 S2 in the Appendix represents the relative percentage of presence for each of the 46 musical features in all improvisations in C2, and Table 6 provides a description of those features that were present in at least 50% of the improvisations in C2. Based both on these results and on acoustic analyses of 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 to 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 more strongly associated with musical feature "01b" (heterometric meter), 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. Approximately 13.9% of the improvisations were grouped in C3 (n=10). Panel C of Figure S2 in the Appendix represents the relative percentage of presence for each of the

Table 5. Musical features of Cluster 2. At least 50% of improvisations contained these features in descending order of importance (100% → 50% present).

Order	Label	Musical Feature	Percentage
1	**08e**	isotonal	100%
2	**01d**	isometric meter	97%
3	**02a**	number of beats is duple	97%
4	**03c**	iso-divisive sub-beats	97%
5	**04a**	simple sub-beats	88%
6	**51**	little or no syncopation	84%
7	**163**	highly repetitive phrases	84%
8	**132**	medium melodic range	81%
9	**191**	highly symmetrical phrases	81%
10	**72**	moderate degree of durational variability	75%
11	**111**	non-hemitonic	72%
12	**151**	small number of overall phrases	72%
13	**09e**	major iso-modal	66%
14	**173**	high degree of repetition with variation	63%
15	**122**	medium melodic interval size	56%
16	**103**	large number of pitch classes	53%
17	**63**	highly motivic	50%
18	**14e**	arched melodic contours	50%

46 musical features in all improvisations in C3, and Table 6 provides a description of those features that were present in at least 50% of the improvisations in C3. Based both on these results and on acoustic analyses of the samples, we find that C3 seems to be primarily defined by the use of *heterometric rhythms* (i.e., semi-regular rhythmic patterns that contain multiple successive meters). These improvisations tended to be more challenging to classify rhythmically, as they contained pauses, hesitations, or unusual phrase combinations that made meter difficult to interpret.

Table 6. Musical features of Cluster 3. At least 50% of improvisations contained these features in descending order of importance (100% → 50% present).

Order	Label	Musical Feature	Percentage
1	**122**	medium melodic interval size	100%
2	**132**	medium melodic range	100%
3	**01b**	heterometric meter	90%
4	**08e**	isotonal	90%
5	**171**	no repetition with variation	80%
6	**72**	moderate degree of durational variability	70%
7	**111**	non-hemitonic	70%
8	**152**	medium number of overall phrases	60%
9	**181**	short phrase lengths	60%
10	**09e**	major isotonal	50%
11	**102**	medium number of pitch classes	50%
12	**161**	non-repetitive phrases	50%

Between-Cluster Differences

Overview. The fact that stylistic heterogeneity between the clusters can be explained by only seven of the 46 musical features (Figure 3) suggests that the sample is relatively homogenous overall. In other words, many of the musical features are either shared between the clusters of improvisations or are unimportant for all clusters (see the open-circle-shaped points in the center of the plot that are not close to any one cluster alone). It is common for cluster analyses to have overlapping features between groups, and this is certainly to be expected in a musical analysis where certain musical features are highly prevalent across the sample. However, it is also clear that it is not only these seven features that contribute to the way each cluster sounds. Rather, these seven features are those that are most specific to each cluster. This suggests that the other 39 features are distributed between clusters in a way that this analysis was not sensitive to. In cases like this, an examination of the percentage of samples per cluster that were coded on particular variable categories (musical features) can serve as additional source of information to help interpret the clusters (Tables 4-6).

Metric (C1, C2) vs. hetero-metric (C3). An important distinction in the results is between improvisations that follow "typical Western music ideals" related to rhythm and meter (e.g., mostly common duple or triple meters) and those melodies that do not (e.g., more heterometric). In the top right hand corner of Figure 3, the musical features "01d" (isometric meter), "02a" (number of beats is duple), "03c" (iso-divisive sub-divisions), "04a" (simple sub-beats), and "51" (little or no syncopation) are grouped together. These features are present in both C1 and C2 to a similar degree, while 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 meters. 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 semi-regular and that contain multiple successive meters (Savage et al., 2012).

Repetition/variation (C2) vs. no repetition/variation (C1, C3). Another important distinction is between those improvisations that contain musical material that is repeated and varied over the course of the melody, compared to 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 Results

Overview. Two one-way ANOVA analyses were carried out to test, individually, whether a) creativity ratings or b) performance ratings could be significantly predicted from the music test scores, if the ratings differed significantly between the improvisation styles, or if there was a significant interaction between the predictive relationship of music test scores between clusters. Before conducting these analyses, we confirmed that the music test scores were a reasonably accurate measure of musicianship when compared with participants' self-reports of musicianship status ("Do you consider yourself to be a musician, yes or no?"). An independent samples t-test was conducted to test whether the difference between the means of the music test score for participants who reported themselves as musicians ($M=7.70$, $SD = 3.32$) or non-musicians ($M=1.82$, $SD = 2.69$) was significantly different from zero. Results showed that musicians scored significantly higher than non-musicians on this test ($t(61.50) = 8.15$, $p < 0.05$, CI [4.43, 7.31], two-tailed).

Creativity (Model A). A one-way between-subjects ANOVA 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, or if there was an interaction between music test scores and musical style. Results showed a significant interaction between membership and music test scores ($F(2,66) = 4.14$, $p = 0.020$, $\eta p^2 = 0.11$), which suggests that the effect of music test score as a predictor of creativity ratings differed between clusters. The effect of music test score was also significant ($F(1,66) = 8.44$, $p = 0.005$, $\eta p^2 = 0.11$). However, interpreting this result would violate the principle of marginality due to the significant interaction, and so it was not interpreted. The effect of cluster membership was not significant ($F(1,66) = 2.37$, $p = 0.10$, $\eta p^2 = 0.067$), 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 ($M=-059$; $SD=0.85$). Figure 4A shows a plot describing this analysis, including the 95% confidence intervals.

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, $\beta = 0.14$, $t(30) = 4.33$, $p = 0.00016$. 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, $r(30) = 0.62$, $p = 0.000082$. The effect of music test scores was not significant for C1, $\beta = -0.0023$, $t(28) = 0.063$, $p = 0.95$, $r^2 = 0.00014$, or C3, $\beta = -0.076$, $t(8) = 0.063$, $p = 0.46$, $r^2 = 0.070$. The correlations between music test scores and creativity ratings for improvisations in C1 ($r(28) = -0.012$, $p = 0.52$) and C3 ($r(8) = 0.26$, $p = 0.23$) were not significant.

Performance quality (Model B). A one-way between-subjects ANOVA was conducted to test if there was a significant difference in performance ratings between musical styles (cluster membership), if music test scores could significantly predict performance ratings, or if there was an interaction between music test scores and musical style. Results showed no significant interaction between music test scores and cluster membership ($F(2,66) = 0.40$, $p = 0.67$, $\eta p^2 = 0.012$. Additionally, there was no significant main effect of cluster membership ($F(2,66) = 1.17$, $p = 0.32$, $\eta p^2 = 0.034$, 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 ($F(1,66) = 7.98$, $p = 0.006$, $\eta p^2 = 0.011$, which suggests there is a significant positive relationship between music test scores and performance ratings. See Figure 4B for a plot describing this regression analysis including the 95% confidence intervals.

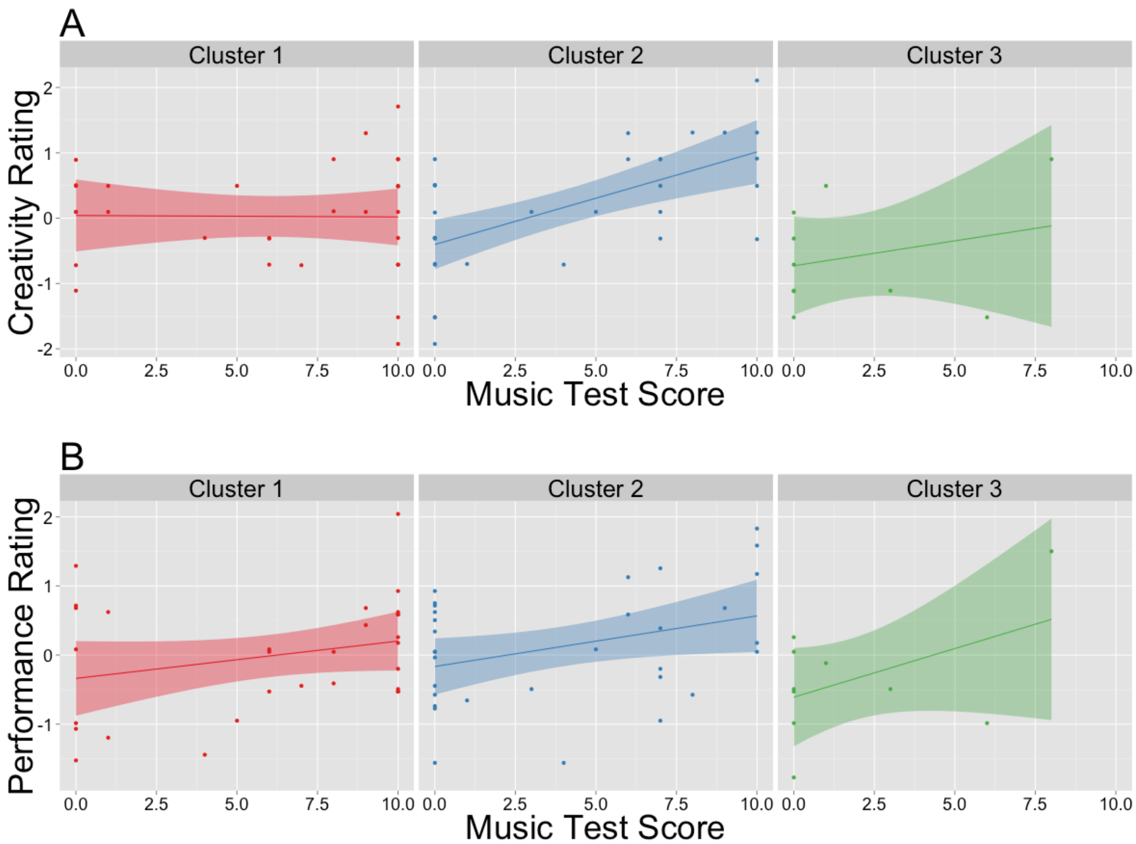

Figure 4. Data from two one-way ANOVA analyses examining the influence of music test scores, cluster membership, and the interaction between music test scores, and cluster membership on ratings of creativity (A) and performance (B). Panel A shows a significant interaction between music test scores and 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 color coding for cluster membership is as in Figures 3 and 4.

CHAPTER 3: General Discussion

The principal objective of this research was to shed light on both the products and processes of musical improvisation by applying classification and clustering methods to a corpus of vocal melodic improvisations produced by novice improvisers in an attempt to uncover relatively discrete clusters of improvisations and to characterize each cluster as a stylistic type ("cantogroup") based on its unique musical features. This represents not only a new approach to analyzing improvisation, but an approach for analyzing creative products in general. 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, while the approach to creativity in the present study was not experimental, our classification/clustering method can be combined with experimental manipulations of the creative task (Fidlon, 2011; Goldman, 2013) so as to examine stylistic changes that result from experimental manipulations. Such approaches permit a more direct inference of mechanisms than is possible without such manipulations. However, even in the absence of such experimental control, a post-hoc analysis of the clusters based on the structural features of the products that are revealed in the combined MCA and k-means cluster analysis permits reasonable inferences to be made about the generative processes (see below).

Our three-cluster solution was able to effect a significant data reduction of the original 72-sample corpus, 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 more-complex and irregular metric structures. Secondly, 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 this cluster, the level of perceived creativity of an improvisation increased as the improviser's music test score increased.

Insights for Cognitive Mechanisms

Based on an analysis of the structural features of the improvisation clusters, we can offer some speculations regarding the cognitive mechanisms underlying the process of improvisation. 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 seem 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 cognitive skills used by expert improvisers (Norgaard, 2011), and 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., pre-learned motor programming) to manage the "note-to-note" choices (Berkowitz, 2010).

Novice improvisers would likely be less able to rely on implicit motor programming for note-to-note choices, and would instead devote much of their working memory to these note-level, rather than phrase-level, decisions. This creates strong limitations for their improvisations because these improvisers may find it difficult to think ahead (Fidlon, 2011). Improvisations in C1 and C3 seemed to demonstrate these characteristics. For example, although improvisations in C1 resembled typical Western melodies, as did improvisations in C2, there was very little phrase repetition and thus no repetition with variation. This might suggest that, although these samples sounded like competent musical melodies, they contained more phrase-to-phase diversity. 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, thereby limiting their improvisations' overall structural organization. Such a focus on note-to-note choices may be especially prominent for improvisations in C3, which appear to be constrained in terms of rhythm, perhaps reflecting hesitations and/or an inability to create a global plan for an improvisation.

Such results can be looked at in the context of fundamental theories of musical improvisation. A pioneering model of jazz improvisation was developed by Pressing (1988, 1998). It suggests that improvisation depends on acquired knowledge, and that improvisers insert pre-learned musical structures at stylistically appropriate times. As the improvisation occurs, the improviser uses perceptual feedback and error correction to monitor and adjust differences between intention and actualization (Pressing, 1988, 1998). In contrast to this, Johnson-Laird (2002) put forth the idea that improvisers do not necessarily have a repertoire of memorized pre-learned patterns, but that they instead generate improvisations based on rule-based procedures derived from the principles of tonal music. We suspect that such a model is going to be more applicable to the kinds of novice improvisers found in our participant pool, who would not have a sufficient training at improvisation to be able to have an internal repertoire of pre-learned musical structures of the kind that jazz musicians typically have. It might be the case, therefore, that

Pressing's is a better model of expert improvisation and that Johnson-Laird's is a better model of novice improvisation.

What Makes a Vocal Improvisation Creative?

Results from the ANOVA analyses showed that there were no significant differences between the improvisation clusters with regard to music test scores, performance ratings, or creativity ratings. However, there was a significant interaction such that music test scores predicted creativity ratings for improvisations in C2. Given that our data demonstrated that 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.

Future work should aim to shed light on the factors that make improvisations creative, while taking advantage of the stylistic analysis made possible by using the combination of classification methods, MCA, and cluster analyses. Such work would benefit from the ability to look at the influence of experimentally-controlled cognitive variables on improvisation style, including working memory (Baddeley, 2003), divergent thinking (Baer, 1996), executive control (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001), domain-specific skills like musical training (Müllensiefen, Gingras, Musil, & Stewart, 2014; Ollen, 2006), and quantity/quality of improvisation experience. For example, (Beaty et al., 2013) showed that divergent-thinking scores for a sample of students of jazz performance significantly predicted the creativity ratings of their improvisations. In future, the inclusion of cognitive manipulations and measurements of musical training could complement our classification and clustering approach for examining improvisational style. The analytical approach we used is ideal for experiments about improvisation that include experimental manipulation. Using cognitive manipulations in improvisation experiments may demonstrate the differential use of musical structures depending on the experimental condition. For example, jazz pianists produce different types of improvisations when they improvise in familiar versus unfamiliar keys (Goldman, 2013).

Additional Considerations

Novices vs. experts. 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. While Mace and Ward's (2002) seminal study of professional visual artists recommended using professional subjects for the study of creativity, due to 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 way (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), as due to a lifetime of passive exposure to music. Singing is a universal human activity that is considered to be one of the most natural means of human expression (Lomax, 1968; Nettl, 1983). Therefore, due to the lack of research on vocal improvisation, we feel that melodic improvisation by novice improvisers should be a topic to be further analyzed by researchers in the field of creativity.

Children vs. adults. 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; Moorhead & Pond, 1978; Moog, 1976; Raju & Ross, 2012; Sundin, 1998; Young, 2002). For example, Moog (1976) discussed how three to four year olds sing spontaneously in ways that involve narrative, imagination, and imitation. Young (2002) suggested that children between one and two years of age are relatively silent in terms of musical improvisation, while three to four year olds sing spontaneously and prolifically. Young (2002) also suggested that after the age of four – when social interactions are more prominent among peers – spontaneous improvisational singing decreases and learned songs become more common. Since spontaneous singing occurs as a natural human behavior, it is an important topic for improvisation research, not least with regard to its structural features. For example, we observed that the improvisations in our sample had an overwhelming tendency to mimic the structure of Western melodies. Since many of our participants did not have musical training, they seem to have demonstrated *implicit* *learning* of complex musical structures as a result of passive exposure to music across their lifetimes, a process referred to in psychology as statistical learning (Saffran, Johnson, Aslin, & Newport, 1999).

Even though singing begins naturally as a process of vocal play during the early years of life, not all individuals continue singing or making music after childhood (Nordoff & Robbins, 1983). When examining everyday creativity, it would perhaps be useful to apply the current approach to children's improvisations. For example, in Raju & Ross, (2012), children were asked to freely generate an original song using the same component of the AIRS Test battery used in the current thesis (Cohen et al., 2009). Using an approach that included intensive classification, the authors employed two descriptive guidelines to organize the compositions into different groups: 1) how well the improvisations fit within the Western tonal musical tradition, and 2) whether or not principles of creating music were applied explicitly or implicitly. This analysis, which was strongly guided by listening, showed that improvisations could be grouped into four styles (Raju & Ross, 2012). The current approach can complement work of this type by providing a more quantitative method for analyzing these products, rather than a method based on listening alone.

Studying novice musical improvisers among adults may be biased towards those participants who are willing to volunteer for a singing experiment due to the possession of certain personal characteristics, such as skill, motivation or commitment with respect to performance (as mentioned by Mace & Ward, 2002). Research on novice improvisers is an important complement to work on expert improvisers. In addition, it is important to keep in mind that novice improvisers can vary strikingly in their musical training. For example, many professional classical musicians would consider themselves as novices when it comes to improvisation. The pool of novice improvisers might have a huge spread of musical training, from total non-musicians to professional musicians. Hence, musical training and musical skills are important moderating variables to analyze in studies of novice improvisations. Again, some novice improvisers might have comparable levels of training and skill to expert improvisers, but simply have no training at improvisation per se.

Choice of classification features. CantoCore (Savage et al., 2012) was modeled after Lomax's Cantometrics classification scheme (Lomax, 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, since the study of musical structure is the dominant focus when analyzing improvisations. The current study piloted the use of a modified version of CantoCore for analyzing improvisations. Of the 26 features of CantoCore, we used a subset of 17 of them and then added two new features related to phrase repetition and repetition with variation that seemed particularly relevant to our sample. Although the musical features used in the current analysis managed to capture important aspects of musical structure related to improvising, the measures used in CantoCore are not necessarily the optimal ones for analyzing improvisation. For example, musical features associated with performance style (e.g., timbre, tempo, rubato, accent) or emotional expression (e.g., Madura Ward-Steinman, 2008) are also important for musical improvisation. While such features are far less reliable to code than structural features (Savage et al., 2012), they could in theory be incorporated into future classification schemes.

The future of research on musical improvisation would benefit from the development and validation of new classification schemes specifically tailored to the analysis of musical improvisation. Some combination of the elements contained in existing schemes for vocal jazz improvisation (e.g., Madura, 1996) and cross-cultural analyses of traditional (folk, indigenous) singing (e.g., Lomax, 1968; Savage et al., 2012) would be a good place to start. Additionally, 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). Finally, the approach of using classification and clustering methods for the analysis of creative products is applicable across the arts, including dance, acting, writing, and visual art. Cross-arts research should be able disambiguate domain-specific and domain-general mechanisms of artistic creativity using similar methods.

Discreteness of clusters. The method of Hwang et al. (2006) demonstrates 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 for cluster interpretation when using this approach in a relatively homogeneous dataset like ours is that each of the 72 samples can only be assigned to a single cluster (as based on the similarities of their CantoCore codings). This presents some drawbacks when defining the clusters based on their musical features, because there are no clear criteria for deciding on which musical features are associated most strongly with each cluster except for visual inspection of the MCA plot. For example, the analysis of the current corpus revealed that at least 39 of the 46 musical features were shared among all three clusters, while only 7 of the 46 strongly differentiated the clusters. This made it challenging to assign a particular feature to a particular cluster in order to describe that cluster's musical style. In other words, many musical features were spread out across the MCA plot Figure 3 and showed no clear proximity to a given cluster centroid. This is a reflection of the homogenous nature of the corpus, rather than a deficiency in the method.

An alternative approach of combining MCA and k-means cluster analysis is the method outlined in (Hwang & Dillon, 2010). This method is a two-way approach to simultaneous MCA and k-means cluster analysis. It is called a two-way approach because it not only assigns each individual improvisation to one specific cluster, but it also assigns each musical feature to a single cluster. Taking the current dataset as an example, the results of a two-way approach would reveal three relatively discrete improvisation clusters, as with the current method by (Hwang et al., 2006), but it would also assign each of the 46 musical features to only one of the three clusters, based on the importance of the features to each cluster. The result would be three clusters of improvisations characterized by unique musical features that do not overlap between clusters. This approach may aid in the stylistic interpretation of the results. However, it would only be advisable in cases with few overlapping features, which was not the case with the current dataset.

Conclusion

Meyer (1989) 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).

The results of the current analysis provide support for the classification-and-cluster approach to analyzing musical improvisation. Three steps are involved in the analysis: 1) classifying the individual products of improvisation using a multivariate classification scheme; 2) generating clusters that describe different stylistic classes of products within the corpus; and 3) testing for associations between these style-clusters and measures of cognition, training, and perceived creativity. This approach offers new insight into the cognitive mechanisms underlying musical improvisation, one that improves upon previous research. Specifically, this approach provides a more holistic and quantitative picture of how people improvise than has been possible in previous studies, not least by explaining a significant proportion of the variance in the data across multiple musical features in a corpus of improvisations. We used CantoCore to classify the data, rather than using either musical transcription or concentrated listening by itself. CantoCore is an easy-to-use classification scheme that captures multiple structural elements of music (Lomax 1976; Savage et al., 2012). The coding of samples by ear using this scheme requires only a modest background in music theory (Savage et al., 2012). Using this approach, we were able to reveal relatively discrete style-types within a corpus of improvisations, and assess their relationship with perceived creativity, as rated by musical experts. 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.

References

Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 651-657). Thousand Oaks: Sage.

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43*(5), 997-1013.

Arabie, P., & Hubert, L. (1994). Cluster analysis in marketing research. In R.P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 160–189). Oxford: Blackwell.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839.

Baer, J. (1996). The effects of task-specific divergent thinking training. *Journal of Creative Behavior, 3*(3), 183-187.

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.

Banjerjee, M., Capozzoli, M., McSweeny, L., & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *Canadian Journal Of Statistics, 27*(1), 3-23.

Beaty, R. E. (2015). The neuroscience of musical improvisation. *Neuroscience & Biobehavioral Reviews, 51*, 108-117.

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, 23*(4), 262-268.

Bengtsson, S. L., Csíkszentmihályi, M., & Ullén, F. (2007). Cortical regions involved in the generation of musical structures during improvisation in pianists. *Journal of Cognitive Neuroscience, 19*(5), 830-842.

Benzécri, J. P. (1979). Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. *Les Cahiers de l'Analyse des Données, 4*(3), 377-378.

Berkowitz, A. (2010). *The improvising mind: Cognition and creativity in the musical moment*. Oxford: Oxford University Press.

Berkowitz, A. L., & Ansari, D. (2008). Generation of novel motor sequences: The neural correlates of musical improvisation. *Neuroimage, 41*(2), 535-543.

Berkowitz, A. L., & Ansari, D. (2010). Expertise-related deactivation of the right temporoparietal junction during musical improvisation. *Neuroimage, 49*(1), 712-719.

Biasutti, M., & Frezza, L. (2009). Dimensions of music improvisation. *Creativity Research Journal, 21*(2), 232-242.

Brown, S., Savage, P., Ko, A. M.-S., Stoneking, M., Ko, Y.-C., Loo, J.-H., & Trejaut, J. (2014). Correlations in the population structure of music, genes and language. *Proceedings of the Royal Society B: Biological Sciences, 281*, 20132072.

Campbell, P. S. (1998). The musical cultures of children. *Research Studies in Music Education, 11*(1), 42-51.

Carson, S. (2006). Creativity and mental illness. Invitational panel discussion hosted by

Yale's mind matters monsortium, New Haven, CT., April 19, 2006.

Chang, W. (1983). On using principal components before separating a mixture of two multivariate normal distributions. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 32*(3), 267-275.

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: Oxford University Press.

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.

Collins, D. (2007). Real-time tracking of the creative music composition process. *Digital Creativity, 18*(4), 239-256.

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America, 121*(2), 1182-1189.

Desarbo, W., Jedidi, K., Cool, K., & Schendel, D. (1991). Simultaneous multidimensional unfolding and cluster analysis: An investigation of strategic groups. *Marketing Letters, 2*(2), 129-146.

De Soete, G., & Carroll, J. D. (1994). *K-means clustering in a low-dimensional Euclidean space*. In E. Diday, Y. Lechevallier, M. Schader, P., Bertrand, & B. Burtschu (Eds.), *New approaches in classification and data analysis* (pp. 212-219). Heidelberg: Springer-Verlag.

Everitt, Brian S., & Horton, T. (2014). *A handbook of statistical analyses using R*. 3rd ed. Boca Raton: 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.

Greenacre, M. (2007). *Correspondence analysis in practice*. 2nd ed. Boca Raton: Chapman & Hall/CRC

Goldman, A. (2013). Towards a cognitive-scientific research program for improvisation: Theory and an experiment. *Psychomusicology 23*(4), 210-221.

Green, P. E., & Krieger, A. M. (1995). Alternative approaches to cluster-based market segmentation. *Journal of the Market Research Society, 3,* 1-11.

Hartigan, J. A., & Wong, M. A. (1979). A k-means clustering algorithm. *Journal of the Royal Statistical Society, 28*(1), 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: Psychology Press.

Hwang, H., & Dillon, W. R. (2010). Simultaneous two-way clustering of multiple correspondence analysis. *Multivariate Behavioral Research, 45*(1), 186-208.

Hwang, H., Dillon, W. R., & Takane, Y. (2006). An extension of multiple correspondence analysis for identifying heterogeneous subgroups of respondents.

Psychometrika, 71(1), 161-171.

Järvinen, T. (1995). Tonal hierarchies in jazz improvisation. *Music Perception, 12*(4), 415-437.

Järvinen, T., & Toiviainen, P. (2000). The effect of meter on the use of tones in jazz improvisation. *Musicae Scientiae, 4*(1), 55-74.

Johnson-Laird, P. N. (2002). How jazz musicians improvise. *Music Perception, 19*(3), 415-442.

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*(2), 151-175.

Kaufman, J. C., Baer, J., & Cole, J. C. (2009). Expertise, domains, and the consensual assessment technique. *Journal of Creative Behavior*, *43*(4), 223.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Le Roux, B., & Rouanet, H. (2010). *Multiple correspondence analysis*. Thousand Oaks: Sage.

Limb, C. J., & Braun, A. R. (2008). Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS ONE, 3*(2), e1679.

Lomax, A. C. (1976). *Cantometrics: An approach to the anthropology of music*. Berkeley: University of California, Extension Media Center.

Lomax A. (1968). *Folk song style and culture*. Washington, DC: American Association for the Advancement of Science.

Lopata, J. A. (2014). *Creativity as a mental state: An EEG study of musical improvisation*. (Doctoral dissertation). University of Western Ontario Electronic Thesis and Dissertation Repository.

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(*2), 179-192.

Madura, P. D. (1996). Relationships among vocal jazz improvisation achievement, jazz theory knowledge, imitative ability, musical experience, creativity, and gender. *Journal of Research in Music Education, 44*(3), 252-267.

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*(1), 5-17.

Meyer, L. B. (1989). *Style and music: Theory, history, and ideology*. Chicago: 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*(4), 621-640.

Moog, H. (1976). *The musical experience of the pre-school child*. London: Schott Music.

Moorhead, G. E., & Pond, D. (1978). *Music of young children*. Santa Barbara: 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*(2), e89642.

Nettl, B. (1983). *The study of ethnomusicology: Twenty-nine issues and concepts*. Urbana: University of Illinois Press.

Nordoff, P., & Robbins, C. (1983). *Music therapy in special education*. St-Louis: MMB Music.

Norgaard, M. (2011). Descriptions of improvisational thinking by artist-level jazz musicians. *Journal of Research in Music Education, 59*(2), 109-127.

Norgaard, M. (2014). How jazz musicians improvise: The central role of auditory and motor patterns. *Music Perception, 31*(3), 271-287.

Norgaard, M., Spencer, J., & Montiel, M. (2013). Testing cognitive theories by creating a pattern-based probabilistic algorithm for melody and rhythm in jazz improvisation. *Psychomusicology, 23*(4), 243-254.

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. (2010). The AIRS on-line test battery of singing skills. In Abstracts of the 2010 CSBBCS annual meeting. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Experimentale, 64*(4), 286–329.

Pfleiderer, M., & Frieler, K. (2010). The jazzomat project: Issues and methods for the automatic analysis of jazz improvisations. In R. Bader, C. Neuhaus, & U. Morgenstern (Eds.), *Concepts, experiments, and fieldwork: Studies in systematic musicology and ethnomusicology* (pp. 279-295). Frankfurt: Peter Lang.

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: Clarendon Press.

Pressing, J. (1998). Error correction processes in temporal pattern production. *Journal of Mathematical Psychology, 42*(1), 63-101.

Raju, M., & Ross, J. (2012). Adaption to Estonian children of the protocol for cross-cultural research in singing. *Trames, 16*(2), 125-144.

Richards, R. (2007). Everyday creativity and the arts. *World Futures, 63*(7), 500-525.

Runco, M. A., & Bahleda, M. D. (1986). Implicit theories of artistic, scientific, and everyday creativity. *Journal of Creative Behavior, 20*(2), 92-98.

Rzeszutek, T., Savage, P. E., & Brown, S. (2012). The structure of cross-cultural musical diversity. *Proceedings of the Royal Society B: Biological Sciences, 279*(1733), 1606-1612.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52.

Savage, P. E., & Brown, S. (2014). Mapping music: Cluster analysis of song-type frequencies within and between cultures. *Ethnomusicology 58*(1), 133-155.

Savage, P. E., Merritt, E., Rzeszutek, T., & Brown, S. (2012). CantoCore: A new cross-cultural song classification scheme. *Analytical Approaches to World Music, 2*(1), 87-137.

Sawyer, R. K. (1999). Improvised conversations: Music, collaboration, and development. *Psychology of Music, 27,* 192-216.

Sawyer, R. K. (2000). Improvisation and the creative process: Dewey, Collingwood, and the aesthetics of spontaneity. *Journal of Aesthetics & Art Criticism, 58(*2), 149-161.

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinex, J. L., & 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*(2), 68-85.

Sundin, B. (1997). Musical creativity in childhood: A research project in retrospect. *Research Studies in Music Education, 9*(1), 48-57.

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, 23*(4), 222-235.

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.

Appendix

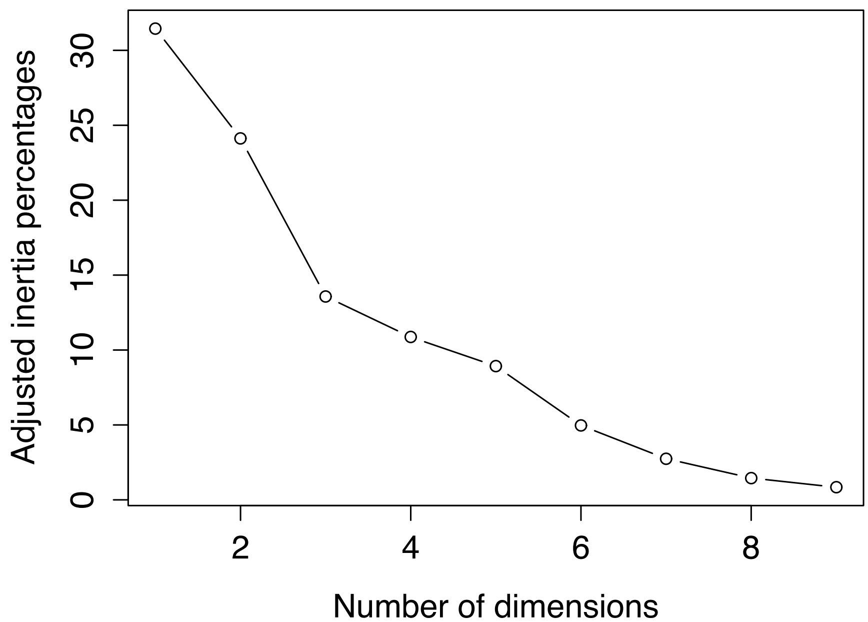

Figure S1. Scree plot for the number of dimensions. The number of dimensions was selected by examining the scree plot and ensuring the parsimonious balance between maximizing the cumulative variance explained by each dimension and minimizing the overall number of dimensions. The optimal number of dimensions for this MCA solution was 3, as defined by the elbow (i.e., transition point) of the scree plot.

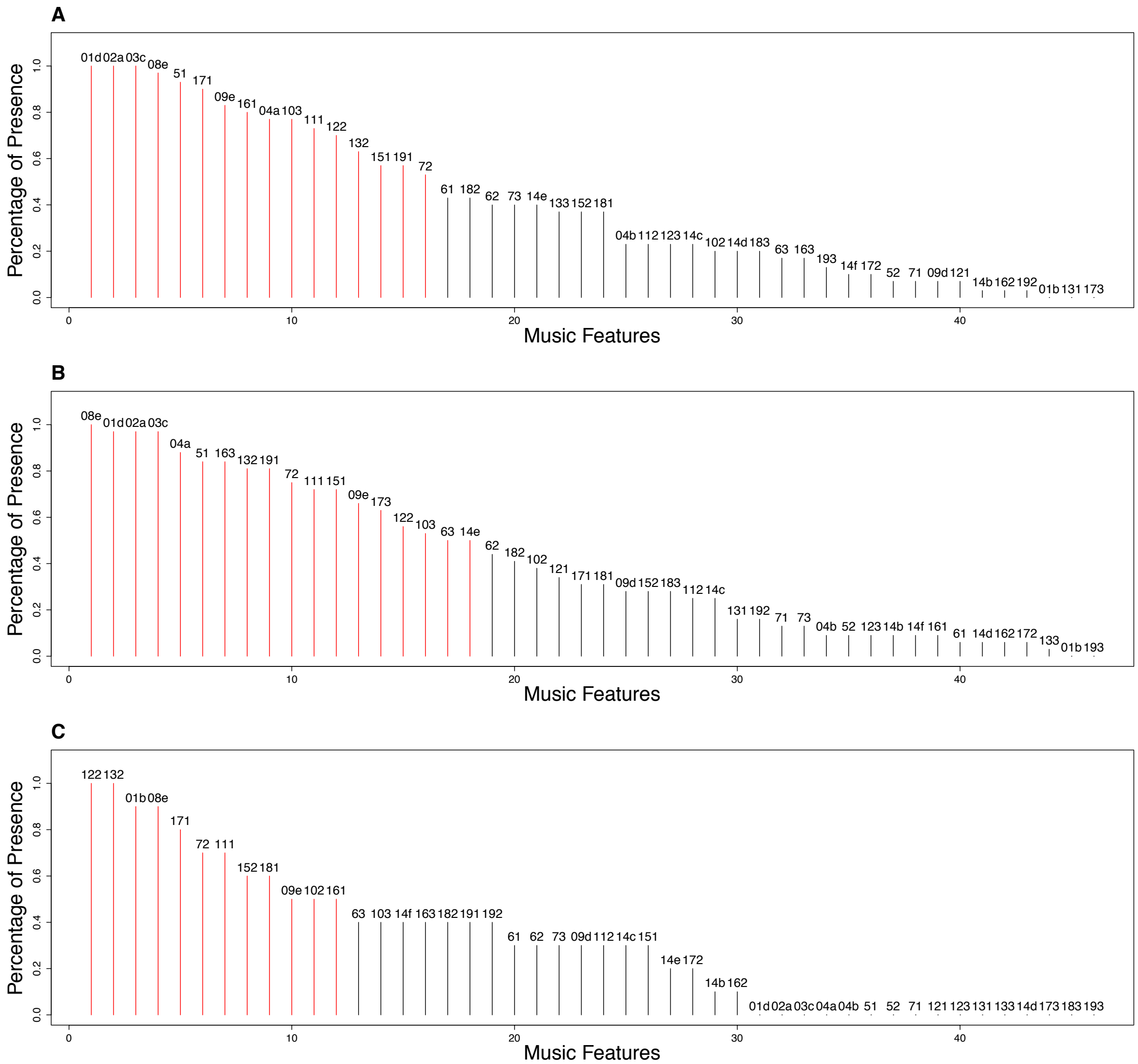

Figure S2. The percentage of "presence" for each of the 46 musical features in the three improvisation clusters. Panel A = cluster 1; panel B = Cluster 2; and panel C = Cluster 3. Red bars indicate those features that are present in 50% or more of the improvisations, and Tables 4-6 outline the definitions of these features.

Table S1. A 10-point music theory test from a music and language background survey created at the University of Prince Edward Island.

Question	Value	Answer
Identify a note on a staff	1	A
Identify a note on a staff	1	F
Identify a chord on a staff	1	Diminished
Identify a chord on a staff	1	Major
Identify a chord on a staff	1	Minor
Identify a musical score	2	*Brother John* (*Frère Jacques*)
Identify a musical score	2	*Twinkle, Twinkle, Little Star*
What does "V-I" refer to?	1	A musical chord progression

Table S2. Modified CantoCore coding scheme.

Number	Structure	Character	Character	Label	Definition
1	Rhythm	Meter	a	01a	A-metric (no consistent beat)
			b	01b	Hetero-metric (consistent beat, but not consistent hierarchical pattern)
			c	01c	Poly-metric (multiple independent beats occur simultaneously)
			d	01d	Iso-metric (there is a single, consistent pattern of strong/weak beats)
2	Rhythm	Number of beats	a	02a	Duple (# of beats can be divided by 2)
			b	02b	Triple (# of beats can be divided by 3)
			c	02c	Complex (# of beats can only be divided by prime numbers higher than 3 (7/4)
3	Rhythm	Beat sub-division	a	03a	A-divisive (beats are not sub-divided)
			b	03b	Herero-divisive (beats are sub-divided, but the number per beat changes)
			c	03c	Iso-divisive (beats are subdivided into consistent # of sub-beats, ex. only eights)
4	Rhythm	Number of sub-beats	a	04a	Simple (can be divided by 2, ex. 4/4, 3/4)
			b	04b	Compound (can be divided by 3, ex. 6/8)
			c	04c	Complex (only divided by prime # over 3)

5	Rhythm	Syncopation (% of notes prominent in metrically unaccented positions)	i	051	Little or no (< 5%)
			ii	052	Moderately syncopated (5-20%)
			iii	053	Highly syncopated (> 20%)
6	Rhythm	Motivic redundancy (% of notes constructed from a single recurring rhythmic pattern)	i	061	Non-motivic (< 20%)
			ii	062	Moderately motivic (20-50%)
			iii	063	Highly motivic (> 50%)
7	Rhythm	Durational variability (# of different types of duration values in a song)	i	071	Low durational variability (less than 3 kinds)
			ii	072	Moderate durational variability (3-4 different kinds)
			iii	073	High durational variability (more than 4 kinds)
8	Pitch	Tonality	a	08a	Indeterminate a-tonal (no discrete pitches, like speech)
			b	08b	Discrete a-tonal (discrete pitches, but no tonic)
			c	08c	Hetero-tonal (tonic modulates/shifts between phrases)
			d	08d	Poly-tonal (multiple, simultaneous tonics in different vocal parts)
			e	08e	Iso-tonal (single tonic throughout)

9	Pitch	Mode (presence of pitch classes at minor 3rd or major 3rd)	a	09a	A-modal: no 3rd present
			b	09b	Hetero-modal: both major and minor 3rd appear in separate phrases
			c	09c	Poly-modal: both appear in the same phrased
			d	09d	Minor iso-modal: minor 3rd only
			e	09e	Major iso-modal: major 3rd only
10	Pitch	Number of pitch classes (# of pitch classes found in the scale)	i	101	Sparse scale <4 pitch classes
			ii	102	Moderately dense scale 4-5 pitch classes
			iii	103	Dense scale >5 pitch classes
11	Pitch	Hemitonicity (% of melodic intervals that are semitones)	i	111	Anhemitonic (<5%)
			ii	112	Moderately hemitonic (5-20%)
			iii	113	Highly hemitonic (>20%)
12	Pitch	Melodic interval size (between successive notes within any vocal part)	i	121	Small intervals minor: 3rd or less
			ii	122	Medium intervals: major 3rd – perfect 5th
			iii	123	Large intervals: minor 6th or greater

13	Pitch	Melodic range (max pitch distance between highest and lowest notes)	i	131	Small range: perfect 5th or less
			ii	132	Medium range: perfect 5th – octave
			iii	133	Large range: more than an octave
14	Pitch	Melodic contour	a	14a	Horizontal: no ascending or descending interval
			b	14b	Ascending only
			c	14c	Descending only
			d	14d	U-shaped: first descending and then ascending
			e	14e	Arched: first ascending and then descending
			f	14f	Undulating: multiple changes of interval direction
15	Form	Total number of phrases	i	151	1-2 phrases
			ii	152	3-6 phrases
			iii	153	> 6 phrases
16	Form	Phrase repetition	i	161	Non-repetitive (through-composed)
			ii	162	Moderately repetitive (3-8 phrases)
			iii	163	Repetitive (1-2 phrases)
17	Form	Repetition with variation (proportion of phrases that are variations on phrases)	i	171	< 24%
			ii	172	25-49%
			iii	173	> 49%

18	Form	Phrase length (phrase length in seconds)	i	181	Short phrases (< 5s)
			ii	182	Medium-length phrases (5-9s)
			iii	183	Long phrases (> 9s)
19	Form	Phrase symmetry (ratio of length of the longest phrase to shortest phrase)	i	191	Symmetric (< 1.5 times the length of the shortest phrase)
			ii	192	Mildly asymmetric (1.5-2.5 times the length the shortest phrase)
			iii	193	Very asymmetric (> 2.5 times the length of the shortest phrase)

Table S3. Pearson product-moment correlations between raters for ratings of creativity and performance.

Creativity Raters	Pearson's r	p-value	Size
R1 vs. R2	0.61	7.89E-09	large
R1 vs. R3	0.27	0.0098	small
R2 vs. R3	0.21	0.039	small

Performance Raters	Pearson's r	p-value	Size
R4 vs. R5	0.63	1.53E-09	large
R4 vs. R6	0.59	2.12E-08	large
R6 vs. R7	0.50	4.33E-06	large