New Ideas in Psychology 57 (2020) 100757

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: http://www.elsevier.com/locate/newideapsych

Bach, Mozart, and Beethoven: Sorting piano excerpts based on perceived similarity using DiSTATIS

Rachna Raman [a,*], Michael A. Kriegsman [a], Hervé Abdi [a], Barbara Tillmann [b], W. Jay Dowling [a]

[a] The University of Texas at Dallas, USA
[b] Lyon Neuroscience Research Center, Lyon, France

ARTICLE INFO

Keywords:
Constrained sorting
Free sorting
Musical style
Musical training
Style perception

ABSTRACT

Our initial aim in this study was to show that Western listeners can sort the music of 3 Western composers consistently on the basis of their compositional style. We found that they could, and proceeded to investigate what cues they might be using to accomplish that task, as well as whether their use of those cues was related to their level of musical training. In Experiment 1, we presented 21 excerpts from the keyboard music of Bach, Mozart, and Beethoven, each excerpt linked to an icon on the computer screen. Participants were to place the icons in different groups following the rule that the icons in one group could have been written by the same composer. First, they did a free sort in which they could form as many groups as they liked, and then we told them that there were just 3 composers, and they should make 3 groups in a constrained sort. In Experiment 1, the excerpts were produced with MIDI transcriptions of the scores, such that the composer's pitch and time information of the notes was preserved, but there was no variation in tempo, dynamics (loudness), or articulation (connectedness or separateness of notes in time). In spite of this simplification, listeners succeeded in clearly differentiating the composers in the constrained sort. In Experiment 2, we used more natural stimuli, 36 excerpts taken from recordings of the 3 composers by 4 pianists who had recorded substantial amounts of each: Arrau, Barenboim, Pirès, and Richter. Here, the stimuli included all the expressive cues of a live performance, and in the constrained sort listeners were even better at categorizing the composers, with not very much difference between the categorizations of trained and untrained listeners. Their judgments were also strongly influenced by the pianists. Richter's performances of the 3 composers were clustered relatively close to the Mozart region of the solution, indicating their clarity and balance; in contrast, those of Barenboim were clustered in the Beethoven region, indicating their sumptuousness and passion. We used a relatively new approach to data analysis—DiSTATIS—which provided the possibility of projecting the sorting results viewed from various perspectives—composer, pianist, participant expertise—into the same space, giving a clearer picture of the results than a piecemeal account of those perspectives.

1. Musical style and its perception

Research on the cognitive processing of musical style is now common in the field of music perception and cognition (e.g., Atalay & Placek, 1997; Crump, 2002; Dalla Bella & Peretz, 2005; Eastlund, 1992; Eerola, Järvinen, Louhivuori, & Toiviainen, 2001; Gardner, 1973; Gingras, Lagrandeur-Ponce, Giordano, & McAdams, 2011; Gromko, 1993; Hargreaves & North, 1999; Miller, 1979; Storino, Dalmonte, & Baroni, 2007; Tekman & Hortaçsu, 2002; Thorisson, 1998; Tyler, 1946; Wedin, 1969; Zivic, Shifres, & Cecchi, 2013). However, earlier research has not yet provided satisfactory answers about the processes underlying style perception. For instance, what features of musical style do listeners perceive in order to categorize the excerpts? Musical style is a complex concept for which a wide range of descriptions has been proposed. Musical style has been defined as a "distinguishing and ordering concept" that "groups examples of music according to similarities between them" (Pascall, 1980). Cope (1991) later defined musical style as "the identifiable characteristics of a composer's music which are recognizably similar from one work to another." Dalla Bella and Peretz (2005) more recently described musical style as that which could refer to a particular musical language (e.g., tonal vs. atonal), to the music composed within a particular historical era (e.g., baroque vs. classical),

* Corresponding author. 800 W. Campbell Road, Richardson, TX, 75080-3021, USA.
E-mail address: rachna.raman@utdallas.edu (R. Raman).

https://doi.org/10.1016/j.newideapsych.2019.100757
Received 31 December 2018; Received in revised form 22 July 2019; Accepted 30 September 2019
Available online 24 October 2019

or to a particular composer's technique (e.g., Mozart's vs. Beethoven's). The defining characteristics of style include recurring phrases, specific forms, melodic, harmonic, or rhythmic features, timbre, typical textures, and formal organization (Meyer, 1973; Vignal, 1987). For instance, Gundlach's (1932) study on the analysis of native American music emphasized the importance of rhythm in categorizing music based on style.

Gardner (1973) defined style recognition as "a complex cognitive process which demands monitoring of numerous aspects of a stimulus, avoidance of over-emphasis upon a single facet, and attention to the general expressive features of the work." Musical style perception can be seen as one example of sophisticated implicit learning processes that lead to nonverbal knowledge just by mere exposure to individual items. Comparable to implicit learning of language or tonal music, listeners seem to become sensitive to structural regularities underlying style (Agres, Abdallah, & Pearce, 2018; Raman & Dowling, 2016, 2017). Although listeners clearly differ in expertise in doing musical tasks related to musical style, such as identifying musical style based on composer or performer, they also share a lot of commonalities in music perception (e.g., Bigand & Poulin-Charronnat, 2006; Gingras et al., 2011; Raman & Dowling, 2017). Daily activities further suggest that listeners, both trained and untrained, are highly sophisticated in recognizing musical styles. For instance, when turning on the radio and listening to music, listeners show remarkable consistency in guessing its musical style (e.g., classical, country, or rock music) or even evaluate finer stylistic differences (e.g., baroque, classical, or romantic periods). And listeners can even easily describe musical excerpts of various styles using similar adjectives (e.g., Hevner, 1936; Miller, 1979; Tekman & Hortaçsu, 2002; Watt & Ash, 1998) and open-ended written descriptions (e.g., Dibben, 2001; Morrison & Yeh, 1999; Thorisson, 1998). Despite the apparent ease of perceptual classification, determining how listeners make such nuanced judgements of style is elusive. As Crump (2002) surmised, an important issue here is to determine the low-level perceptual and high-level abstract information (i.e., cues) that listeners perceive and respond to so as to make judgements of musical style. As early as Hevner (1936), studies have shown that some of the primary features that listeners seem to focus on in understanding a musical piece include mode (major vs. minor), harmony, and rhythm. Musical style, and its perception, thus creates a puzzle: Listeners easily and rapidly recognize the style of an historical period or of a composer, but researchers have been largely unsuccessful at providing theoretical descriptions that capture the characteristics of musical style and its perception.

Musical style perceptions are also influenced by listeners' music training and general music listening experiences. Meyer (1989) stated that perception of style seems to be a learned behavior for musically trained listeners, which is the case since understanding the chronological development of various musical styles is a necessary part of their music education. For instance, Miller (1979) found that untrained participants tended to focus more on affective qualities of the music (e.g., playful, carefree, pleasing, etc.), whereas expert participants focused more on musical structure and other technical aspects (e.g., form, dynamics, harmony, etc.). Miller also identified an historical dimension but only for the experts, who were able to categorize styles mostly through differences in composers' use of harmony. On the contrary, several other studies have shown that music novices (e.g., Thorisson, 1998; Tyler, 1946; Wedin, 1969) and nonmusicians (e.g., Eastlund, 1992; Gingras et al., 2011; Gromko, 1993), and even non-Western nonmusicians are sensitive to style recognition (e.g., Dalla Bella & Peretz, 2005), and that perhaps such sensitivity is developed merely due to prolonged exposure to any type of music and the perception of low-level musical cues.

In understanding listeners' style perceptions, one particular area of interest is how listeners categorize musical pieces based on stylistic aspects into various genres, such as folk, classical, popular, and so forth. Most of the studies in the field have investigated style perception either via ethnomusicology, wherein researchers quantitatively analyze the prototypical melodies of a particular style of music by studying the statistical distribution of different intervals, pitches, or temporal patterns (e.g., Gundlach, 1932; Zivic, Shifres, & Cecci, 2013), or via the use of trained computer networks and simulations to imitate human performance (e.g., Järvinen, Toiviainen, & Louhivuori, 1999; Smaill & Westhead, 1993). Some researchers have even compared computer simulations versus human performance on style identification and categorization (e.g., Atalay & Placek, 1997; Crump, 2002; Tillmann, Abdi, & Dowling, 2004). Through such analyses, researchers have systematically examined both low-level perceptual characteristics, such as statistical patterns of notes, and high-level abstract features, such as harmony, rhythm, and melody, of music that contribute towards style perception.

Studies using sorting tasks, which are implicit and do not require verbalization of stylistic features by human listeners, are relatively rare, mostly due to the difficulty in finding suitable methods to accurately measure listeners' perception of style. Earlier studies mostly relied on tasks where participants had to explicitly verbalize their judgements of style (e.g., Gardner, 1973; Tyler, 1946). One of the earliest studies on style perception was conducted by Tyler (1946), who had novice music students listen to 3-min excerpts of three selections each from Mozart, Beethoven, and Schubert. These excerpts were from different movements within the same piece. The stimuli were presented randomly and the students had to verbally judge who the composer of each selection was. The study was performed twice by the same participants on two different occasions during the semester. The results indicated that on both occasions participants showed sensitivity to style recognition, and this sensitivity was related to their prior music training and concert experience, but not with their intelligence scores or their preference of the composers or music pieces.

Gardner (1973) conducted a developmental investigation of style, and the obtained data from a verbalization task showcased the importance of participant age in musical style perception. Five groups of participants (ages 6, 8, 11, 14, and 18–19 years) judged whether pairs of 15-s excerpts of classical music from 1680 to 1960 belonged to the same piece or not. Participants 11 years and older performed similarly and more accurately than the younger participants, who also showed some sensitivity to such discriminations. Also, participants 11 years and older could discriminate pieces based on subtler aspects of style, and were able to categorize the pieces as either from the same musical era or from different eras.

With advancement in statistical methods, researchers began using nonverbal "rating" tasks, such as the similarity-judgement task, and multidimensional scaling (MDS) to analyze subjective ratings, thus popularizing its application to investigate recognition of musical style. The similarity-judgement task involves nonverbal categorization of musical stimuli via rating responses. For instance, Wedin (1969) investigated the perceptual dimensions into which the historical epochs of music would fall using MDS. Participants heard 98 pairs of 10-s excerpts (including repetitions) of Western classical music from 1720 to 1890, and rated them in two types of similarity-judgement tasks. In the first task, novices with less than 5 years of music training judged the degree of similarity in percentage between pairs of excerpts. In the second task, a similar procedure was followed and three participant groups (highly and moderately trained groups, and novices) had to rate on a 10-point scale the subjective similarity in style of the pair of excerpts. The results revealed that all participants showed style sensitivity but participants with greater musical training grouped the musical excerpts into four distinct clusters—Baroque, Rococo, Viennese Classicism, and Romanticism. Thus, they showed a clearer and nuanced distinction among the categories. On the contrary, participants with lesser or no musical training grouped the musical excerpts into three distinct clusters—Baroque, Viennese Classicism, and Romanticism.

Eastlund (1992; also Gromko, 1993) further extended this approach using the same nonverbal task to investigate differences in style perception among untrained participants, music undergraduates, and experts (music professors). She used 15 musical pieces belonging to either the classical or the romantic styles, composed between 1762 and 1896. Participants heard 105 pairings of 15-s music excerpts and performed a similarity-judgement task using a 7-point scale. MDS analysis showed that music undergraduates and expert musicians performed almost identically, and their combined responses could be classified into three dimensions (in order of importance): historical period in which the piece was composed, perceived complexity of the excerpt, and its tempo. On the contrary, for untrained participants historical period was the least important, which partially explains Miller's (1979) findings of a lack of historical dimension for untrained participants. Eastlund proposed an explanation for this difference in style perception, that untrained participants may focus more on what she called secondary features of music (e.g., tempo, pitch, dynamics, etc.), whereas trained participants focus more on primary features (e.g., melody, harmony, etc.).

Later, Thorisson (1998) examined the validity of style-categorization results from the nonverbal similarity-judgement task by comparing with participants' open-ended written descriptions. He examined whether novice listeners were able to classify musical excerpts as either classical or romantic, based on compositional styles. Participants were first familiarized with 17 classical and romantic piano excerpts, and then they completed similarity ratings of the 136 possible pairings of the excerpts. MDS indicated that the excerpts were generally grouped into two clusters, one for the classical period and the other for the romantic era. Listeners gave written descriptions of attributes pertaining to texture, tempo, dynamics, and so forth, for each piece, and the results showed that excerpts from the same musical period but by different composers received similar attributes, thus validating the use of both tasks.

Several studies have used the nonverbal similarity-judgement task to determine the exact nature of the musical cues listeners employ to identify the genre of a musical piece. For example, Eerola et al. (2001) used similarity ratings of folk melodies to predict music students' classification of melodies that represented five different European folk styles. MDS analysis showed that the students were able to categorize the melodies based on the different folk styles. The results also indicated that the salient aspects of a musical piece, such as statistical properties of the pitches and rhythm, to which listeners generally paid attention, presumably helped listeners classify the pieces according to melodic similarity.

Supplementing Eerola et al.'s (2001) findings, Tekman and Hortaçsu (2002) used a verbalization task to determine how listeners perceived the relationship among various musical styles. They asked Turkish undergraduate students to list all the genres of music they knew and had them rate a list of adjectives on how appropriately they described each genre. MDS identified two components based on the students' classification of the different styles: historical novelty (traditional vs. modern) and appeal (to large population vs. to small groups) of the styles. The students classified closely associated styles, such as rap and techno, systematically. Qualitative analyses showed that the students also described the musical styles on the basis of three dimensions—evaluative, activity, and peacefulness. Tekman and Hortaçsu's study provided evidence that listeners not only possess knowledge on various styles of music but also on the relationship of the styles to each other, and the unique descriptive qualities associated with each style.

Later, Dalla Bella and Peretz (2005) investigated the recognition of musical style using a different approach with the same nonverbal similarity-judgement task. They had two professional composers create 16 piano excerpts imitating the musical styles of baroque, classical, romantic, and post-romantic eras, and six advanced piano students recorded the excerpts. Each student played excerpts from only one style so that confusion between compositional styles and performance styles could be avoided. Since the excerpts were composed specifically for the study, the researchers could control for confounds (e.g., familiarity with the musical piece). Western music students, and Western and non-Western (Chinese) non-music students performed a familiarity-rating task for each excerpt and a similarity-judgement task for 128 excerpt pairs. Half of the excerpt pairs were presented in the historical order (e.g., classical followed by post-romantic) whereas the other half of the pairs were presented in the inverse order (e.g., post-romantic followed by classical). MDS analysis showed that all participants rated melodies from earlier historical periods (e.g., baroque) as more familiar, and they rated compositions closer in styles (i.e., historical eras) as similar. This sensitivity to style recognition was enhanced in the musician group, though both Western and non-Western nonmusician groups also showed an obvious sensitivity, indicating the significance of mere passive long-term exposure to music. Dalla Bella and Peretz also proposed that universal low-level perceptual processes (such as, temporal regularities) may underlie style sensitivity. And finally, the results showed an order effect, wherein participants differentiated the styles more easily when the excerpts were presented in chronological order rather than when reversed.

Storino et al. (2007) further extended Dalla Bella and Peretz's (2005) research by investigating whether familiarization with a single composer's musical grammar can facilitate listeners' style categorization based on that composer's technique and of the corresponding historical period in general. In Experiment 1, expert musicians in the baroque style were first familiarized with eight Legrenzi's (an Italian baroque composer) arias. In the test phase, participants heard excerpts from 10 arias by Legrenzi and 10 arias produced by LEGRE (a computer program that produces new arias in Legrenzi's style for the same texts of music). Half of the participants heard the excerpts as 10 Legrenzi-LEGRE pairs and the remaining participants heard the 20 excerpts in isolation (i.e., not paired). Both groups had to identify which of the excerpts was composed by Legrenzi. All participants were able to classify based on style, however, accuracy was higher in the paired condition and was just above chance level in the isolated condition. In Experiment 2, trained (not in any particular style) and untrained participants performed only the paired task. The results showed that only the trained participants were able to perform above chance level. The method of Experiment 4 was the inverse of Experiment 1, wherein trained participants were first familiarized with eight arias produced by LEGRE. In the test phase, they heard only 18 arias by three Italian baroque composers—Legrenzi, Rossi, and Gabrielli—and not those produced by LEGRE. Participants had to indicate whether the arias had been created by the same composer in the familiarization phase (i.e., LEGRE) or not, and the results showed that participants were successful, thus confirming the similarity between LEGRE and Legrenzi styles of composition. Storino et al. found that with brief exposure, even musicians non-experts in the baroque style (as in Experiments 2 & 4) were able to perceive stylistic features of the music.

Storino et al. (2007) used a sophisticated grammar based on musicological analysis in their LEGRE program. A contrasting approach in recent years relies on simply capturing the sequential regularities of the musical style in the music in a Markov-based statistical learning program (see Agres et al., 2018, for a review). Listeners have been shown to extract the sequential regularities of melodies they were exposed to, and expectancies are fairly well matched by the statistical learning programs. However, as Krumhansl (2015) points out, citing Meyer (1989), there is more to musical style and listeners'—even untrained listeners'—understanding of it than can be captured with Markovian statistical learning. Music has structural and expressive properties that are easily grasped by the attentive listener, but which are not taken into account by Markovian statistics.

To summarize, researchers over the years have used various types of verbal and nonverbal tasks in order to ascertain how listeners perceive and categorize stylistic aspects of music. The results show a broad picture that trained and untrained listeners, and even untrained young children, are generally sensitive to stylistic aspects of music, and expertise enhances the perception of style. However, most of the earlier work has only been able to speculate about the types of cues used by trained and untrained listeners in such tasks. One reason is that the types of tasks used in previous studies may not have been suitable to answer the primary question: How do listeners categorize musical excerpts based on stylistic aspects? Studies involving the measurement of implicit processes, such as those applied in the categorization of stimuli, should use appropriate indirect or implicit investigation methods to obtain the best possible results. The sorting task is designed for measuring implicit processes, such as those involved in most music-related tasks. For instance, Brown (1981) found that trained and untrained participants agreed less with their group performance when they had to pair melodies with descriptive words provided by the researcher (explicit) versus when they did the matching task by providing their own words (implicit). Similarly, Dibben (2001) found differences in participant responses between nonverbal and verbal categorization tasks, wherein participants were more inclined to group two sounds when they resembled each other acoustically in the implicit nonverbal task, whereas they were more inclined to group them by their physical source in the explicit verbal task. A second reason is that earlier tasks involving style perception and its analyses only provided results obtained by averaging group responses, and there was no way to track individual responses. This could be due to the fact that the statistical tools used to measure and analyze the multidimensional aspects typically characteristic of human responses were not sophisticated enough, and did not provide the possibility of projecting the results from various perspectives (e.g., composer, participant expertise) into the same space, thus giving a fragmentary account of those perspectives.

The only study we know of that has used a sorting task to categorize musical style (note that here it was not composer's style but performer's style of playing) was conducted by Gingras et al. (2011). Experts and non-experts heard organ excerpts represented as icons on a computer screen, which they had to sort into six groups based on the performer's playing style. The excerpts were played by three award-winning and three non-award-winning organists, who rendered two versions each of expressive and inexpressive interpretation of the same piece. The results indicated that both trained and untrained participants were able to accurately sort excerpts based on tempo, wherein faster excerpts were differentiated from slower excerpts, and articulation (connectedness or separateness of notes in time), wherein expressive performances were differentiated from inexpressive ones. Also, participants' sorting was influenced by performer competence, in that they accurately differentiated between award-winning versus other performers.

Although Gingras et al. (2011) successfully used the sorting task in a well-controlled setting, in which participants heard versions of the same excerpt played in different ways by six organists, an important issue that should be further studied is whether listeners can categorize excerpts of different composers played by different performers based on stylistic features. Also, Gingras et al. only used the constrained sort, which prompted us to investigate whether participants would be able to categorize by composer's style when they are first told to sort freely into as many groups as they see fit.

2. Goals and Hypotheses

The purpose of our study was to extend previous findings by examining the influence of compositional style, type of sorting task (i.e., free vs. constrained), type of stimuli (MIDI vs. natural), pianists' playing style, and listeners' music expertise on their ability to perceive stylistic aspects in musical excerpts. We used both free and constrained sorting tasks and an updated version of the statistical tool DiSTATIS, which had never before been used in studies pertaining to music perception and cognition. One advantage of the sorting method is that the judgements are more likely to reflect the multiple dimensions of the stimuli than would have been the case when using the earlier paired comparison similarity judgements. Moreover, sorting tasks do not involve any form of verbalization, thus tapping into the listeners' implicit knowledge. And such a nonverbal approach facilitates the use and assessment of stylistic cues. Our study addressed the following questions: (1) Are listeners able to sort brief melodies based on compositional style? (2) If so, does the type of sorting task—free versus constrained— interact with music expertise in influencing listeners' perception? Investigations involving other features of music, such as emotion (e.g., Bigand, Filipic, & Lalitte, 2005; Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005), as well as research not involving music (e.g., Scott & Canter, 1997) have measured listener responses in both sorting tasks while presenting the two tasks sequentially, free sort followed by constrained sort. However, none of the sorting studies pertaining to musical style that we have referenced have tested this, and so we decided to investigate what might prompt differences, if any. Especially, we wanted to examine whether untrained participants in particular could produce coherent categorizations for composer's style with the free sort. (3) Does the type of stimuli—MIDI versus natural—influence the task? In contrast to the previous studies, we compared listeners' perception of stylistic aspects in music between both MIDI and natural stimuli excerpted from commercial recordings. (4) Also, will the performance of four different pianists influence musical style perception? Unlike Dalla Bella and Peretz (2005), we wanted to examine whether individual playing styles will influence participants' sorting choices and the degree to which the different playing styles would affect listeners' perception of the composers. (5) Finally, does the listener's music training influence the perception of stylistic aspects of music?

Based on earlier studies (e.g., Crump, 2002; Dalla Bella & Peretz, 2005; Eastlund, 1992; Eerola et al., 2001; Gardner, 1973; Gromko, 1993; Hargreaves & North, 1999; Miller, 1979; Storino et al., 2007; Tekman & Hortaçsu, 2002; Thorisson, 1998; Tyler, 1946; Wedin, 1969), we hypothesized that there would be an effect of musical period and compositional style, and that listeners would identify greater stylistic differences among pieces from eras farther apart. That is, participants would more distinctly categorize pieces by Bach and Beethoven versus those by Bach and Mozart, or Mozart and Beethoven. Our second hypothesis was that, in general, participants would be faster and more accurate in their perception of style in the constrained sort when compared to the free sort. We based our prediction on the fact that participants completed the constrained sort immediately following the free sort, which made them somewhat more familiar with the excerpts. Also, the instructions were "clearer" with the constrained sort, where we disclosed the actual number of composers. We expect that the results might indicate how far increased familiarity and the change in instruction would change the result pattern. Our third hypothesis was partly based on Gingras et al.'s (2011) findings, that participants would be more accurate at perceiving the stylistic aspects in an expressive performance (natural stimuli) when compared to a more mechanical one, since the natural stimuli have richer dynamics and tonal qualities, thus facilitating perception. Our fourth hypothesis was also partly based on Gingras et al.'s (2011) findings, that in Experiment 2, listeners would be influenced by the performance of the pianists in their sorting of the pieces. Thus, composers' style would interact with pianists' style wherein participants might incorrectly attribute pieces to other composers due to confusion with the pianist's performance style. Our last hypothesis was that highly trained musicians would be more accurate than untrained participants at perceiving the stylistic aspects. Previous research has shown that professional musicians perceive musical structures differently from amateur musicians (Dowling, 1986), with experts performing better than amateurs at a variety of musical tasks (Krampe & Ericsson, 1996). However, several studies have also shown that untrained participants are sensitive to underlying structural and affective patterns of music, and are able to perform several musical tasks above chance levels (e.g., Bigand & Poulin-Charronnat, 2006; Dalla Bella & Peretz, 2005; Eastlund, 1992; Gingras et al., 2011; Gromko, 1993; Wedin, 1969). Thus, we also expect that untrained participants would display some style sensitivity.

3. Framework for compositional style

In this study, we used piano excerpts from three composers: Bach, Mozart, and Beethoven. The composers' compositional styles are classified into three different epochs: Bach's style is classified as baroque, Mozart's as classical, and Beethoven's as romantic. There is general agreement among musicologists that Bach is a prototypical baroque composer who played a very special role in the baroque period (Grout & Palisca, 1980). Nevertheless, his compositions stood out in the baroque era due to their melodic, harmonic, and rhythmic complexity. In general, though Bach occasionally shifted emotional tone in the middle of a movement (especially in his cantatas), he typically followed the baroque style of maintaining a constant emotional tone throughout a movement. Composers in the generation after Bach (e.g., Josef Haydn and Bach's son Carl Philipp Emmanuel), began to experiment with emotional shifts within a movement, techniques that Mozart exploited during the classical period. In Beethoven, the range and frequency of emotional shifts was expanded even further. Beethoven is right at the start of the romantic approach, and his earlier works are usually viewed as transitional between the classical and the romantic.

Especially in Experiment 2, we selected almost all of Beethoven's excerpts from his romantic period. In accordance with the main features of musical style outlined above, we can characterize these three styles in terms of the variability of the musical material along the dimensions of pitch, time, and loudness, as well as of musical texture (dense vs. open and transparent) and timbre. Timbre is not so much an issue in the present experiments because all the excerpts are played on the piano, or in a piano timbre. And variations in loudness and tempo are only an issue in Experiment 2 where we used actual performances of the pieces, in contrast to the MIDI transcriptions of Experiment 1 which do not vary in loudness or in tempo. The harmonic language differs among these styles—the way in which chord progressions and key relationships are handled as the music develops in time (e.g., Zivic et al., 2013). Also, Krumhansl (2015) cites results showing that there are differences in interval patterns between baroque (largely scale wise) and romantic (largely arpeggios) styles. And there is a definite change in the variability of emotional tone within an excerpt as we progress through the early (baroque) and the middle (classical) 18th century and then on to the 19th (romantic).

The baroque style exemplified by Bach is characterized by rhythmic regularity as well as relative stability of loudness, pitch, texture, and emotional tone within an excerpt. The harmonic language is dense and complicated, taking surprising turns which are then resolved to achieve expected ends. The texture typically consists of the interweaving of two or three separate melodic lines in different pitch registers, which are clearly discernable. This is in contrast, for example, to music that consists of a single melodic line accompanied by block chords in which the individual pitches are not distinctly heard.

The classical style emerged from the baroque through the innovations of mid-century composers, such as C.P.E. Bach, Haydn, and Mozart. The harmonic language becomes simpler, often with a slower progression from chord to chord, but shifts of tonal center (modulations) are often more abrupt, and signal a shift in emotional tone. Rhythmic organization also becomes more irregular than in the baroque, accompanied by greater variation in loudness. Textures are more varied, with dense as well as open textures, and often with a single melodic line with chords or repetitive melodic figures outlining chords as accompaniment.

Beethoven, starting to write in the 1790s, shifted music into the romantic style. Here, the tendencies apparent in classical music become accentuated. Especially for Beethoven (in contrast to later romantics, such as Chopin, Schumann, and Brahms), the harmonic language becomes even more simplified. Beethoven is sometimes inclined to emotional outbursts indicated by abrupt changes in loudness, tempo, and texture. The range of pitches typically in use, expanded somewhat in the classical compared with the baroque, is now widely expanded.

One aspect to consider with these three composers is that, spanning a century as they do, their influence on each other is one-directional. Mozart was a dedicated admirer of Bach, and from time to time there are unmistakable signs of Bach's influence. Mozart's String Trios, K. 404a, consists of his arrangements of Bach preludes and fugues, along with additional pieces to go with them that he wrote in the same style. And the duet for the Two Armed Men in *The Magic Flute* is definitely written in the style of Bach, which gives it a seriousness and solemnity important to that scene in the opera. And Beethoven drew on both Bach and Mozart in his piano music and string quartets, in which he includes passages where the interweaving of simultaneous melodic lines is reminiscent of Bach. Beethoven was fond of Mozart's piano concertos, especially Concerto No. 20 in d minor, which he often played in concerts and for which he wrote a cadenza.

4. Framework for pianist style

Whereas in Experiment 1 we used MIDI transcriptions that simply reproduced the notes on the printed page with no stylistically induced nuances in performance, such as variations in loudness, tempo, and phrasing, in Experiment 2 we used excerpts from commercial recordings played by four pianists: Claudio Arrau, Daniel Barenboim, Maria-João Pirès, and Sviatislav Richter. We picked these pianists because they were among the relatively few pianists in the middle and late 20th century who had recorded substantial amounts of the works of the three composers. (Many pianists are known for concentrating on one composer, or several composers in a similar stylistic period. Arthur Rubenstein, known for Beethoven and the later romantics, for example, rarely if ever recorded Bach or Mozart, and Rosalyn Tureck, a Bach specialist, rarely recorded Mozart or Beethoven.) We also selected them because their personal styles of playing the piano differed systematically. Richter is widely regarded as presenting each composer, and each piece, in its own terms, without imposing a particular personal imprint, but with considerable emotional engagement (Villemin, 1999). His Bach is transparent and lucid, in that the inner melodic lines are rarely obscured, but it is also forceful, as are his Mozart and Beethoven. Arrau has been described as leaving his own imprint on the pieces he engages with (Villemin, 1999), but his Bach is also transparent and his emotional engagement is very clear. In contrast to Richter's playing which can often strike one as jagged and craggy, Arrau's is much smoother, but equally sensitive to the emotional tension. Pirès has a much lighter touch than either Richter or Arrau—transparent and lucid with all three composers, and often more playful. Barenboim—and here we are talking of the Barenboim of the 1960s and 70s, and not the mature Barenboim evident in his recordings during the last 10 years—definitely leaves his own imprint on all the pieces, and it is an imprint most suited to the highly emotional Beethoven. He uses the sustain pedal of the piano much more than the other pianists, which aids in the buildup of emotional climaxes, but inhibits transparency in open textures, such as those of Bach and Mozart.

In our two experiments, undergraduate and graduate students from an American university participated in the study. All participants reported having normal hearing and a regular school education of at least 12 years. We obtained informed consent from each participant before the start of the experiment, and all participants completed a brief questionnaire on their musical experience. Participants included musicians at various levels of training (as measured by years of formal training).

5. Experiment 1: MIDI stimuli

5.1. Method

Participants. Thirty-nine participants, mean age 22.47 years (range = 18–29 years), took part in Experiment 1. Eleven participants reported that they had no music training whereas the remaining 28 participants reported that they had between 1 and 30 years ($M = 7.54$

years) of formal music training.

Stimuli. Stimuli consisted of 21 excerpts from seven keyboard pieces each by Bach, Mozart, and Beethoven (see Appendix A). We selected the excerpts for each composer from plain MIDI transcriptions available on the Internet. All excerpts were of pieces written for piano or harpsichord, and we avoided pieces we judged to be relatively familiar, such as those found in elementary piano books, like Bach's short minuets, Mozart's *Sonata in C, K. 545*, or Beethoven's *Für Elise*. We produced .wav files of CD quality from MIDI files, as follows: The MIDI files had been transcribed directly into MIDI form from the musical scores, with no attention to nuances of dynamics, phrasing, or variations in tempo (e.g., files found at http://www.madore.org/~david/music/midi/). The excerpts were converted into .wav files using Cakewalk Professional version 4.0 using an acoustic piano voice. Each excerpt lasted for 9–10 s ($M = 9.67$ s). We linked the excerpts to audio icons arranged pseudorandomly on a PowerPoint slide. Participants listened to the excerpts over good quality headphones.

Sorting tasks: Free and constrained. A sorting task is a simple but useful method to examine implicit nonverbal processes, such as listeners' perception of stylistic aspects of the melodies. Sorting tasks can reveal the underlying structure of a collection of items, in this case musical excerpts. Similar to similarity-judgement and rating tasks, sorting tasks access implicitly learned knowledge, in this case knowledge about stylistic aspects of melodies. However, sorting tasks are considered to be more effective than judgement and rating tasks as they are less strenuous on the participants, and can be used to compare experts and non-experts without relying on either a specialized vocabulary or a quantitative response (Chollet, Valentin, & Abdi, 2014). For instance, we conducted interviews with participants in the pilot study about their experiences with the sorting task, and they all uniformly reported that the task was fun and not at all tiring. Some researchers have indicated that untrained participants and especially children have difficulty in verbalizing their perceptual responses to art forms (e.g., Gardner, 1973). Other researchers have shown that participants' performance change in a verbal versus nonverbal musical task (e.g., Brown, 1981; Dibben, 2001). An added theoretical advantage for a nonverbal sorting task could be that participants are able to use their own concepts for categorizing stimuli especially in a free sort (e.g., Scott & Canter, 1997), and are completely in control of the experiment in terms of its pace and time limit, so they probably find the task less taxing. As in Bigand, Vieillard, Madurell, Marozeau, and Dacquet (2005), Bigand, Filipic, and Lalitte (2005), and Scott and Canter (1997), the order of the two sorting tasks, free and constrained, could not be alternated as the purpose of the free sort was to have participants categorize the excerpts without specific direction, whereas the purpose of the constrained sort was to re-categorize the same excerpts based on specific instructions. Sorting tasks are commonly employed in studies on sensory perception, such as food preference and quality. As far as we know, only one study (i.e., Gingras et al., 2011) has employed this method to assess listeners' perception of musical style, though listeners regularly apply sorting methods in their everyday musical and nonmusical activities.

Procedure. In our experiment, we adapted the methodology used in Bigand, Filipic, et al. (2005) and Bigand, Vieillard, et al. (2005), and Bigand, Filipic, et al. (2005). We instructed participants to listen to each excerpt by clicking on its icon, and then to sort the icons into clusters based on their perceived similarity to each other—in particular based on whether they might have been written by the same composer. While sorting, participants could play each excerpt in any order they wished and as many times as they wanted similar to the methodology used by Bigand et al. and Gingras et al. (2011), and especially as the stimuli were presented to them in a random order with each task. However, we did not register the number of times participants heard each stimulus nor the order in which the stimuli were heard. Participants completed two types of sorting tasks sequentially: free sorting and constrained sorting. In the free sorting task, participants sorted the 21 excerpts into as many clusters as they thought necessary, with the constraints that there should be at least two clusters, and that each cluster should contain at least two excerpts. In the constrained sorting task, participants were required to sort the excerpts into three clusters only as we had excerpts from three composers; this gave them more direction and should have helped them in sorting. The whole task took approximately 20–40 min to complete, depending on how often the participant listened to the various excerpts.

Data Analysis. We recoded each participant's sorting data as a distance matrix. Excerpts sorted together were assigned a distance of 0, whereas excerpts sorted into different groups were assigned a distance of 1. To analyze the perceived differences among the excerpts, we then used an updated version of DiSTATIS (Abdi, Williams, Valentin, & Bennani-Dosse, 2012). DiSTATIS is a generalization of two multivariate methods: metric multidimensional scaling (MDS; Abdi, 2007b), a method for analyzing a single distance matrix, and STATIS, a method for executing multi-table principal component analysis (PCA; Abdi et al., 2012; Abdi & Williams, 2010). DiSTATIS is commonly used to assess multiple distance matrices, such as data from sorting tasks (Abdi, 2007a; Abdi, Valentin, Chollet, & Chrea, 2007), wherein each participant produces a distance matrix. Here our application of DiSTATIS relies on *a priori* knowledge, namely the fact that we used excerpts from exactly three composers (and in Experiment 2, four pianists).

In DiSTATIS, participants' distance matrices are double-centered, normalized, integrated (i.e., combined), and decomposed to give a factor map. To double-center the matrices (Abdi, 2007b), a distance matrix is converted to a covariance matrix centered on the origin. In this way, double-centering brings disparate matrices to the same center (similar to centering as in calculating z scores). Double-centered matrices are normalized in the style of multiple factor analysis (Abdi, Williams, & Valentin, 2013), where each double-centered matrix is divided by its first eigenvalue so that the scales of the tables are comparable. These double-centered and normalized tables are then subjected to an analysis of between-table similarity, called R_V analysis (Abdi, 2010), in order to identify typical and atypical tables. The R_V analysis provides a set of table weights, such that atypical tables receive small weights. The weighted average of these tables gives the best possible single representation of all the tables, called the compromise table (Abdi et al., 2012). Finally, the compromise table is decomposed by PCA to generate components. Thus, DiSTATIS reveals the best possible single representation of the perceived relationships among the stimuli.

The advantage of using DiSTATIS is that, unlike MDS and PCA, it retains the information provided by the pattern of each participant's responses, but like MDS and PCA, DiSTATIS produces new variables, called components (also called dimensions, factors, or principal axes). Components are ordered by strength and are mutually orthogonal. That is, the first component explains the maximum possible variance, and the subsequent components explain the maximum remaining variance under the constraint that each subsequent component is orthogonal to all prior components. The coordinates of the stimuli on the components are called factor scores.

For ease of visualization, typically two components are plotted in what is called a component map. On this map, observations are interpreted by their distances from each other and their positions on the components. Observations near each other are similar. An observation that has a large factor score on a given component contributes much variance to that component. Each component may reflect an effect measured along that dimension, which may relate to a perceived difference between the observations (e.g., staccato vs. legato). Thus, two observations on the same side of a component are perceived as similar on that dimension, whereas observations on opposite sides of a component are perceived as different on that dimension. In the figures, dots representing excerpts are color-coded by composer, and a square box in the appropriate color indicates each composer's average position.

We also performed inference tests in the form of nonparametric bootstrap resampling to test the stability of differences between groups. We tested the differences among the three groups of composers, and also among three levels of music training of the participants: 11 untrained participants, 10 moderately trained musicians (1–4 years of training, $M = 2.60$ years), and 18 highly trained musicians (5 or more years of training, $M = 10.28$ years). Previous studies have shown that people with 5 or more years of formal music training perform differently on musical tasks than those with less than 5 years of training or those with no training at all. For example, Dowling (1986), and Dowling and Bartlett (1981) showed strong differences in performance between people with average of 5 years of music lessons than those without any. Bootstrap resampling consists of resampling participants within groups with replacement (DiCiccio & Efron, 1996), a procedure intended to simulate sampling from the population of individuals from which the participants are drawn. Bootstrap samples are collected repeatedly (here, 1000 times) to generate successive distributions of the groups. For each group, the most extreme 5% bootstrap-sampled means are removed, leaving a peeled convex hull that contains 95% of the bootstrap-sampled means, giving a 95% bootstrap confidence interval. For visualization, a smoothed ellipse is fitted around the convex hull, and so is slightly more conservative than the convex hull itself. We conducted the analyses in R (version 2.15.2; R Core Team, 2012), adapting the DistatisR (Beaton, Chin-Fatt, & Abdi, 2014a, 2014b) and the MExPosition (Chin-Fatt, Beaton, & Abdi, 2013) packages to that use.

6. Results

We conducted DiSTATIS analyses on the data from Experiments 1 and 2 separately for the free and constrained sorting tasks. Table 1 shows the percent of variance explained by the first four components in each of the four overall analyses in which the sorting was based on composers. These components explain between 5.28 and 21.47% of the variance in the four analyses.

6.1. Free sorting

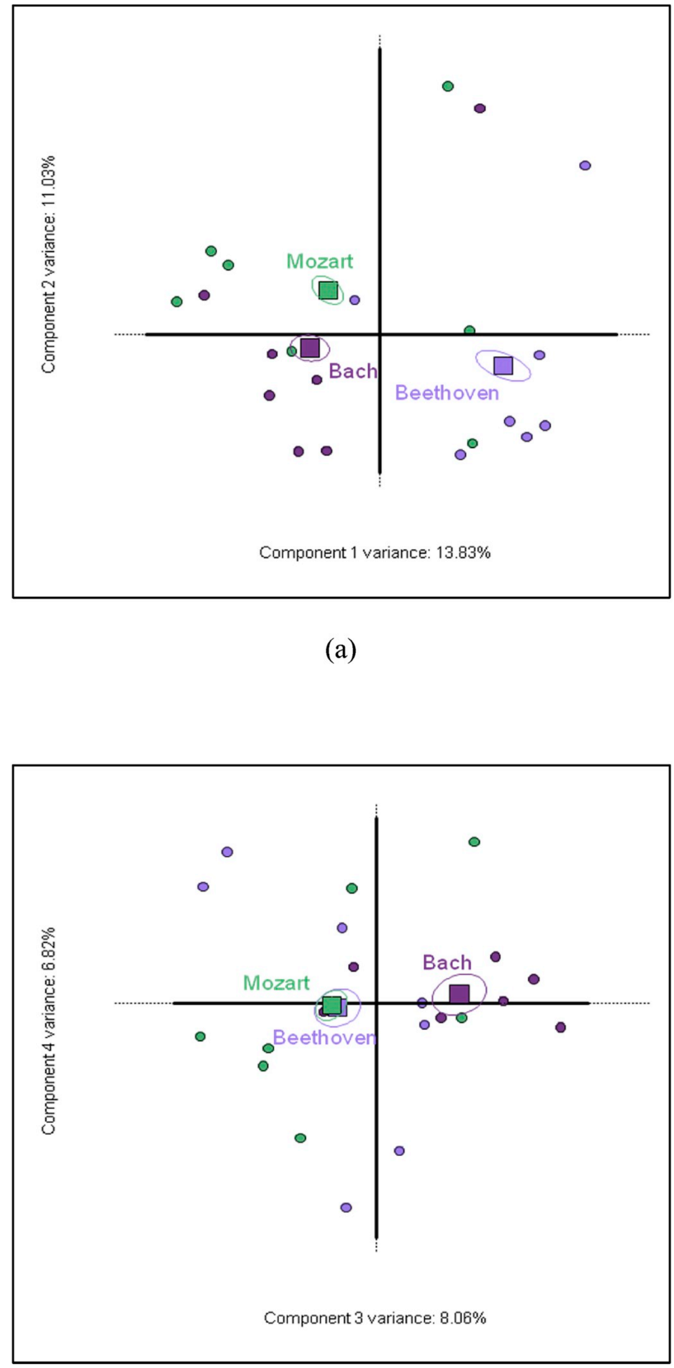
Composers. Fig. 1 shows that Components 1, 2, and 3 captured the effects of composer. Component 1 differentiated Beethoven from the other two composers. To a lesser extent Component 2 differentiated Mozart from the other two. Component 3 differentiated Bach from Mozart and Beethoven, whereas for Component 4 there were no apparent differences due to composer.

Music Training. Fig. 2 shows the results of the R_V analysis for sorting patterns produced by the participants. Here, each dot corresponds to a participant. Participants were color-coded according to level of music training: highly trained, moderately trained, and untrained. Component 1 displayed a non-significant effect in which highly trained musicians were separated from the others. Component 2 indicated a separation between moderately trained musicians and untrained participants, but with highly trained musicians in between. Subsequent components did not reveal between-group effects.

6.2. Constrained sorting

Composers. In Fig. 3, Components 1, 2, and 4 showed that pieces by Beethoven were clearly distinguished from those of the other composers. Component 3 differentiated Bach from the other composers.

Music Training. Fig. 4 shows the results of the R_V analysis. No effects of music training were found; and the three groups behaved more similarly with constrained sorting than they did with free sorting.



(a)

(b)

Fig. 1. Compromise factor scores for Experiment 1: Free sorting task with MIDI stimuli, color-coded by composers (Bach: purple; Mozart: green; Beethoven: lavender). Panel (a) Components 1 and 2, and Panel (b) Components 3 and 4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Variance (%) explained by the first four components in Experiments 1 and 2 in terms of composers.

	Experiment 1		Experiment 2	
	Free	Constrained	Free	Constrained
Component-1	13.83	19.70	10.58	21.47
Component-2	11.03	13.21	7.93	13.03
Component-3	8.06	9.44	5.89	7.03
Component-4	6.82	7.83	5.28	5.76

7. Experiment 2: Natural stimuli

7.1. Method

Participants. Thirty-seven participants, mean age 23.27 years

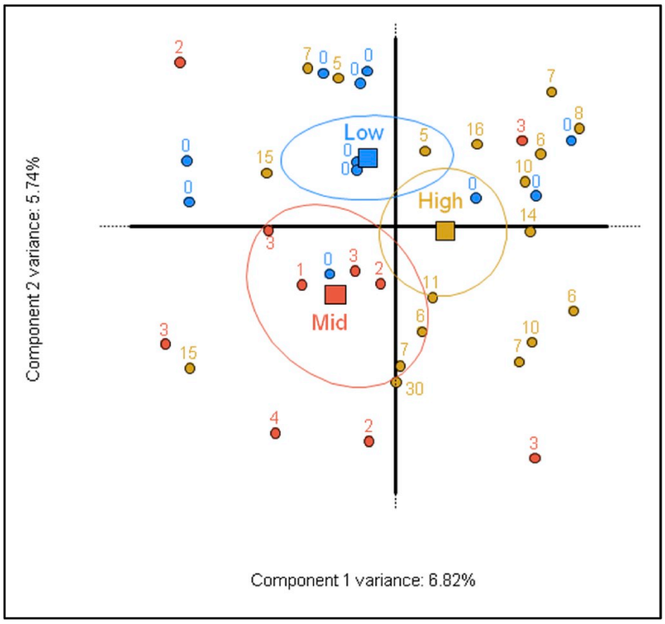


Fig. 2. R_V factor scores for Experiment 1: Free sorting task with MIDI stimuli, color-coded by musical experience (highly trained: orange; moderately trained: red; untrained: blue). Each dot represents a participant and the numbers corresponding to each dot represent the years of music training. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

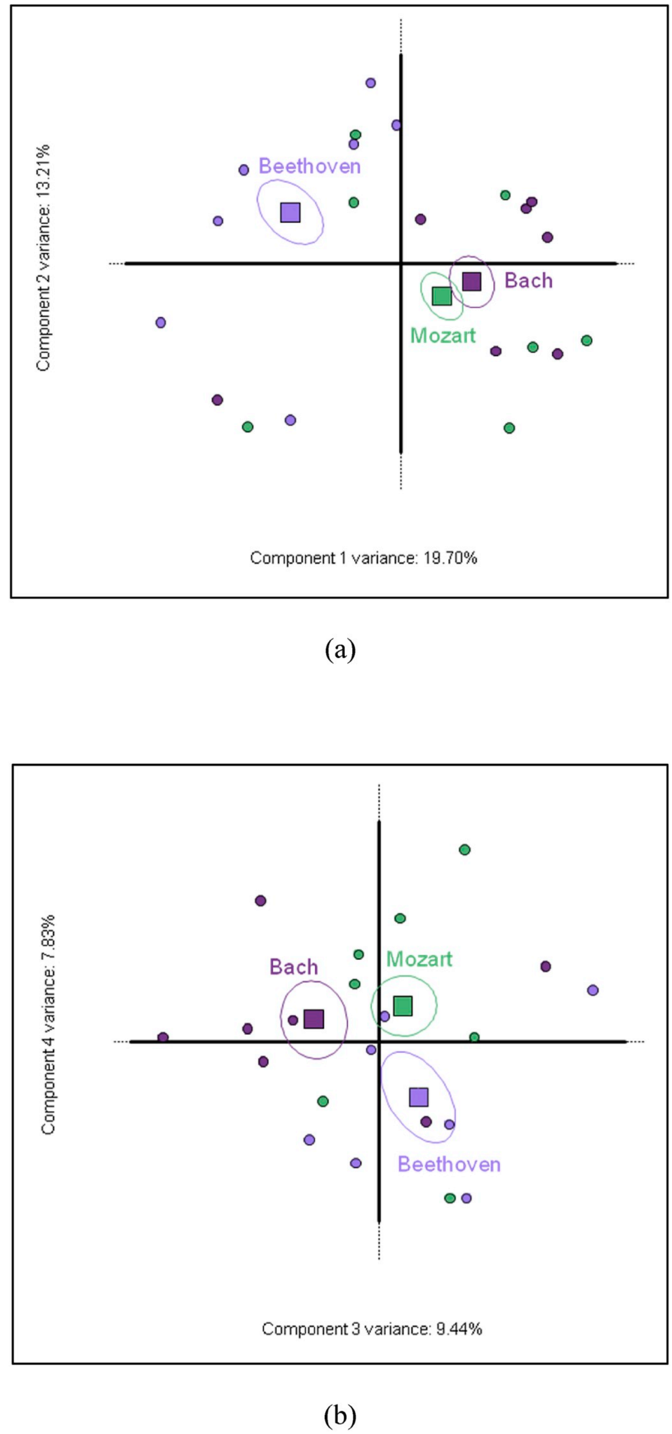


Fig. 3. Compromise factor scores for Experiment 1: Constrained sorting task with MIDI stimuli, color-coded by composers (Bach: purple; Mozart: green; Beethoven: lavender). Panel (a) Components 1 and 2, and Panel (b) Components 3 and 4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(range = 17–50 years), took part in Experiment 2. Ten participants reported that they had no music training whereas the remaining 27 participants reported that they had between 1 and 15 years ($M = 4.89$ years) of formal music training.

Stimuli. Stimuli consisted of 36 newly selected excerpts from commercial CD recordings: 12 pieces each by Bach, Mozart, and Beethoven. Each of four pianists—Arrau, Barenboim, Pirès, and Richter—played three different pieces by each composer (see Appendix B). This enabled us to assess the constancy of a composer's place in the sorting patterns across varied pianists, and the degree to which differences among the pianists affected sorting. As in Experiment 1, we avoided relatively familiar works. We were constrained by the selection of works that the particular pianists had recorded. For example, Richter had mainly recorded Bach's *Wohltemperierte Klavier*, whereas Arrau had mainly recorded partitas and suites. In contrast to Experiment 1, these excerpts exhibited all the nuances of phrasing and dynamics characteristic of musical performances. Each excerpt was at least 9 s in length, and ended at a musically coherent place, so that they varied in length from 9 to 15 s ($M = 11.64$ s). Presentation of stimuli was the same as in Experiment 1.

Procedure and Data Analysis. The procedure and data analysis were the same as those of Experiment 1 except for the following differences: The total duration of the task was approximately 30–45 min, depending on the participant. And the groupings addressed by the DiSTATIS nonparametric bootstrap resampling analyses included contrasts among the four pianists as well as among composers. Categorization of participants' expertise was the same as in Experiment 1, with 10 untrained participants, 17 moderately trained musicians ($M = 2.06$ years), and 10 highly trained musicians ($M = 9.70$ years).

8. Results

8.1. Free sorting

Composers. In Fig. 5, Components 1 and 2 differentiated Mozart from Beethoven, with Bach's excerpts clustered near the origin. Component 3 differentiated Bach from Beethoven, and Component 4 differentiated Mozart from the other two.

Pianists. Fig. 6 shows the results regarding pianists. Component 1 differentiated Richter from Pirès. Component 2 differentiated Richter and Pirès from Barenboim, with Arrau in the middle. Component 3 differentiated between Richter and Barenboim with Arrau and Pirès in the middle. Arrau was consistently positioned near the origin. Note that Figs. 5a and 6a suggest a connection between composer and pianist, such that Barenboim appears to be definitely associated with Beethoven,

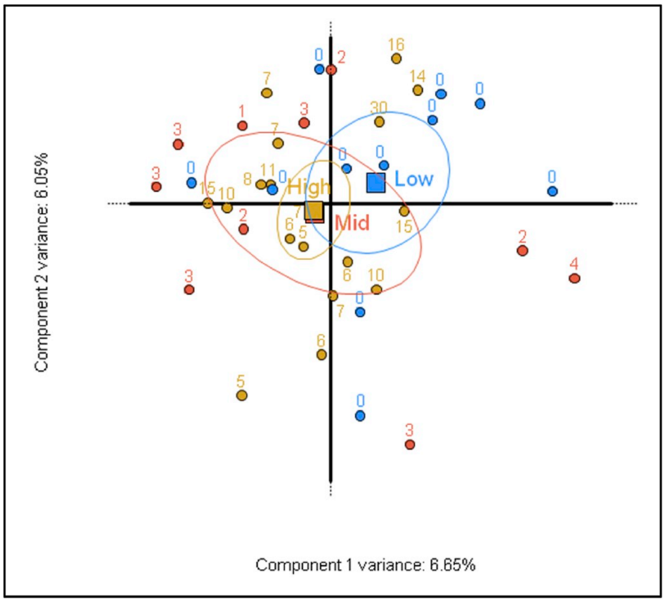


Fig. 4. R_V factor scores for Experiment 1: Constrained sorting task with MIDI stimuli, color-coded by musical experience (highly trained: orange; moderately trained: red; untrained: blue). Each dot represents a participant and the numbers corresponding to each dot represent the years of music training. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

with Pirès and Richter more associated with Mozart, and Arrau appears at the origin along with Bach.

Music Training. Fig. 7 shows the results of the R_V analysis. Component 1 indicated an effect of musical experience, significantly separating low and high levels of musical training with moderate levels in between. There were no other clear effects.

8.2. Constrained sorting

Composers. In Fig. 8, Components 1, 2, and 4 distinguished between Beethoven and Mozart, whereas Component 3 differentiated Bach from the other two.

Pianists. Fig. 9 shows the results in terms of pianists. Richter and Barenboim were consistently perceived as distinct. Component 2 differentiated Pirès from Barenboim. Components 3 and 4 taken together distinguished Barenboim and Arrau from Richter and Pirès. Note that Figs. 8a and 9a show a relationship between composer and pianist, similar to that seen in Figs. 5a and 6a.

Music Training. Fig. 10 shows the results of the R_V analysis. Component 1 displayed a non-significant difference indicating a weak effect of musical training. There were no other effects.

9. Discussion

In considering these results, let us first look at the contrast between free sorting and constrained sorting. In general, constrained sorting produced greater agreement among the listeners than free sorting (which was done first), as shown by the amount of variance explained by the successive factors in the DiSTATIS solutions (see Table 1). Especially in Experiment 2, the gain attributable to constrained sorting is substantial. In both experiments, the total amount of variance explained for constrained sorting by the first four factors is around 50%, compared with about 40% in Experiment 1 and about 30% in Experiment 2 for free sorting. Constraining the sorting to just three categories forced listeners to make difficult choices of whether to put excerpts in the same cluster, which they had perhaps avoided in the free sort by creating more categories. And those choices led to greater consistency and agreement

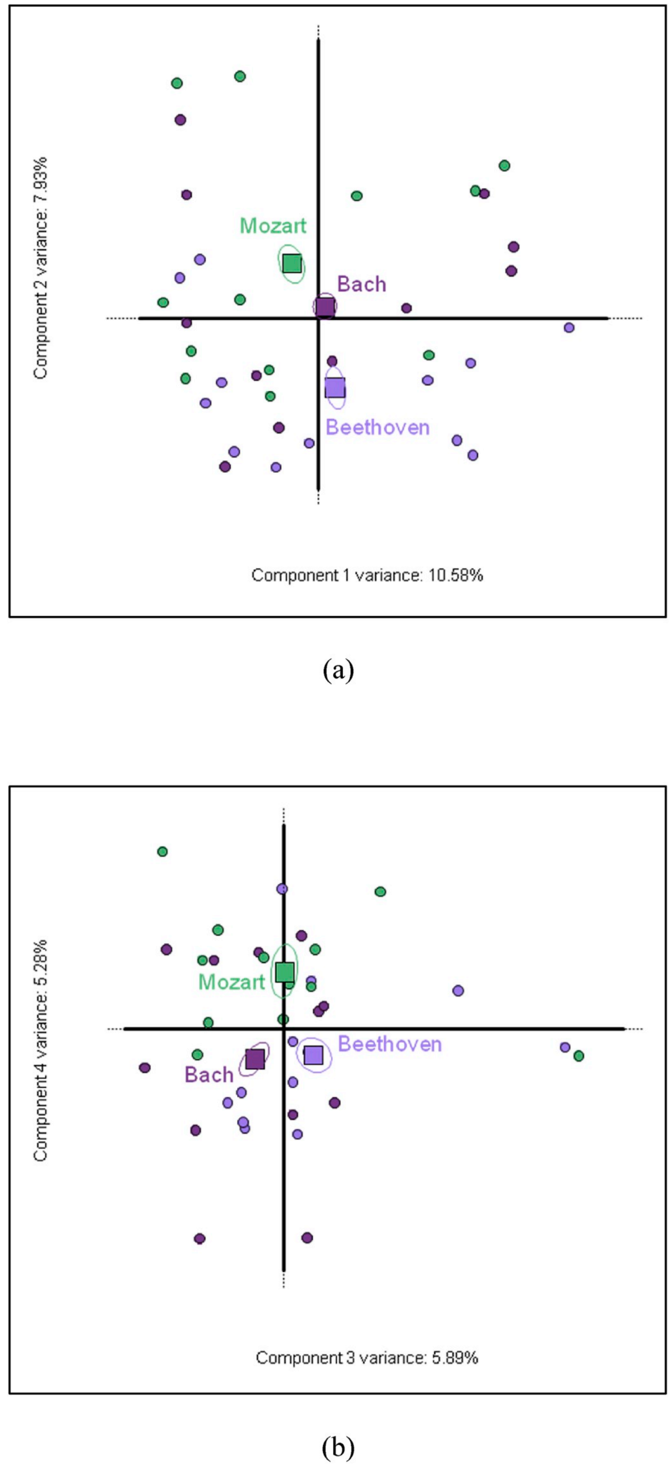
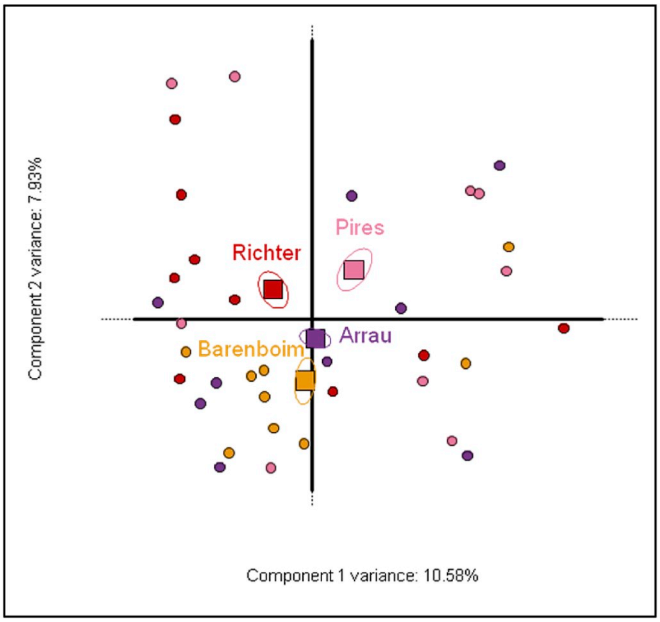
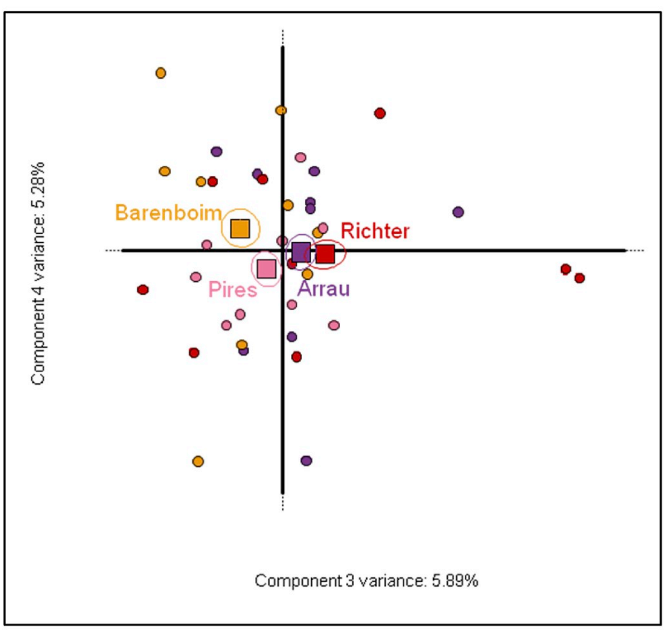


Fig. 5. Compromise factor scores for Experiment 2: Free sorting task with natural stimuli, color-coded by composers (Bach: purple; Mozart: green; Beethoven: lavender). Panel (a) Components 1 and 2, and Panel (b) Components 3 and 4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

among the listeners in their categorization of style. This increase in consistency was accompanied by greater convergence among the groups of listeners with different amounts of musical training, as seen in going from Figs. 2–4 for Experiment 1, and from Figs. 7–10 for Experiment 2. These results suggest that in constrained sorting, the untrained and the moderately trained groups appear to be using much the same features for making decisions about compositional style. And there is



(a)



(b)

Fig. 6. Compromise factor scores for Experiment 2: Free sorting task with natural stimuli, color-coded by pianists (Arrau: purple; Barenboim: orange; Pirès: pink; Richter: red). Panel (a) Components 1 and 2, and Panel (b) Components 3 and 4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

considerable overlap between the features they use and the features used by the more highly trained groups. In general, in this regard these results agree with those of Dalla Bella and Peretz (2005), Eastlund (1992), Gingras et al. (2011), Gromko (1993), and Wedin (1969). These results also concur with those of Brown (1981) and Dibben (2001), in that untrained and trained listeners perform similarly in nonverbal music tasks. In regard to Miller's (1979) finding that untrained listeners tend to rely more on affective qualities of the excerpts, we note that the convergence across training levels was more emphatic in Experiment 2, where those affective qualities were more evident in the naturalistic excerpts, than in Experiment 1, where they were largely absent.

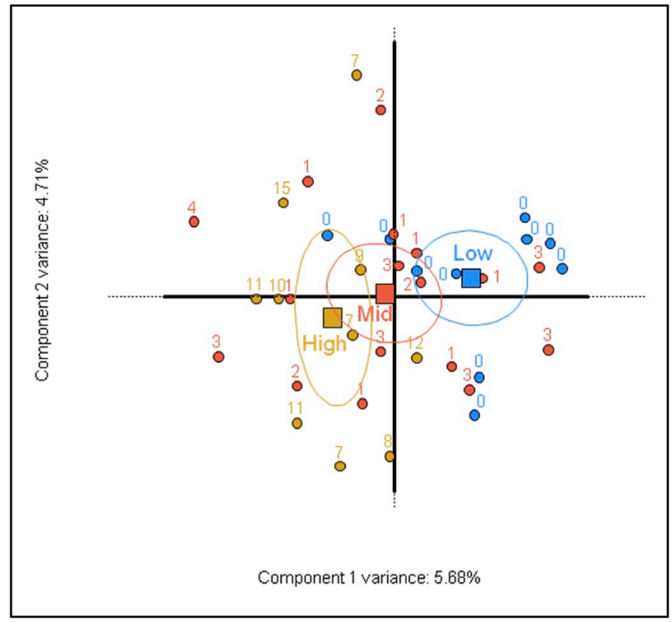


Fig. 7. R_V factor scores for Experiment 2: Free sorting task with natural stimuli, color-coded by musical experience (highly trained: orange; moderately trained: red; untrained: blue). Each dot represents a participant and the numbers corresponding to each dot represent the years of music training. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Since the constrained sorts were more coherent than the free sorts, in what follows we will concentrate on them. In Experiment 1, Fig. 3a shows that the first component tends to separate the three composers according to their historical order: Bach, Mozart, and Beethoven. This is in agreement with the results of Dalla Bella and Peretz (2005) who suggested that this categorization was largely driven by the increase in rhythmic freedom as style developed from the baroque through the classical to the romantic. Such an increase in rhythmic freedom involves features that would be quite evident in the MIDI versions of Experiment 1, so this interpretation strikes us as entirely appropriate. The second component in Fig. 3a appears to contrast Bach and Mozart with Beethoven. Among the readily available features in the MIDI excerpts, harmonic complexity suggests itself as underlying this contrast: Bach and Mozart are notably more complex in their harmonic progressions than Beethoven, especially the relatively early Beethoven represented in Experiment 1 (see Appendix A). (This local trend runs counter to the more general historical trend noted by Dalla Bella and Peretz of an increasingly freer use of the tonal system over the last three centuries.) The third component (Fig. 3b) contrasts Mozart and Beethoven with Bach, and may have to do with constancy of texture. As noted above, Bach's writing typically involved the simultaneous presentation of two or three separate melodic lines in a texture that remains generally constant throughout an excerpt, and this textural consistency is obvious in these MIDI excerpts. Mozart and Beethoven, in contrast, shift their textures often, between few versus many simultaneous notes, and between pitch regions, and those shifts are also obvious in the MIDI transcriptions. Component 4 (Fig. 3b) again contrasts Bach and Mozart with Beethoven, but we do not venture an interpretation.

Turning to the naturalistic excerpts of Experiment 2, we see in Component 1 (Fig. 8a) an ordering of Beethoven-Bach-Mozart. With the live pianists we think this reflects differences in the forcefulness of the

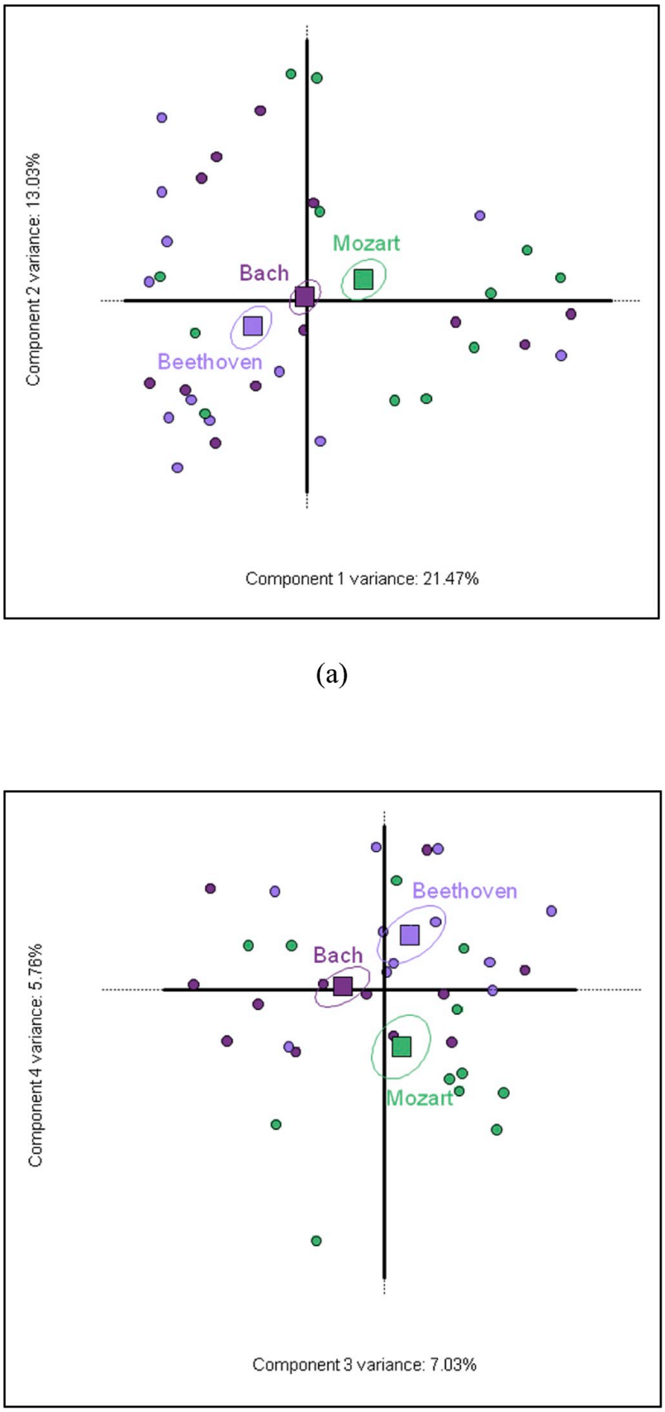


Fig. 8. Compromise factor scores for Experiment 2: Constrained sorting task with natural stimuli, color-coded by composers (Bach: purple; Mozart: green; Beethoven: lavender). Panel (a) Components 1 and 2, and Panel (b) Components 3 and 4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

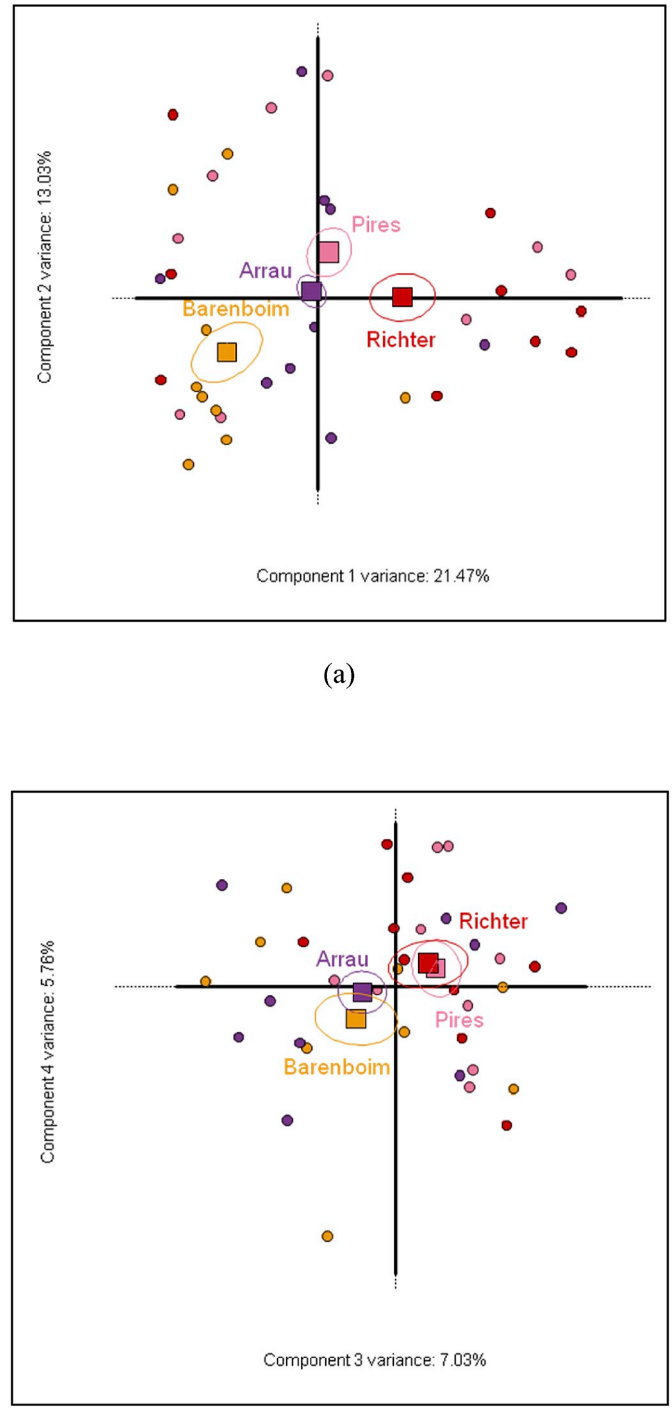


Fig. 9. Compromise factor scores for Experiment 2: Constrained sorting task with natural stimuli, color-coded by pianists (Arrau: purple; Barenboim: orange; Pirès: pink; Richter: red). Panel (a) Components 1 and 2, and Panel (b) Components 3 and 4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

performances, involving dynamic (loudness) contrasts. Beethoven uses the greatest dynamic contrasts, and with these pianists Bach comes a close second, whereas Mozart is more reserved and delicate. Component 2 (Fig. 8a) appears to reflect large-scale rhythmic unpredictability, in which the less predictable Beethoven is contrasted with the more predictable Bach and Mozart. This contrast was accentuated in the live performances because the pianists tended to give dynamic emphasis to Beethoven's rhythmic surprises, which led to a different result here than in Experiment 1 (see Component 1 in Fig. 3a) where no such emphasis could occur. As a result, the three composers do not line up in historical order on what we are thinking of as a dimension of rhythmic complexity as they did in Experiment 1 and in Dalla Bella and Peretz (2005).

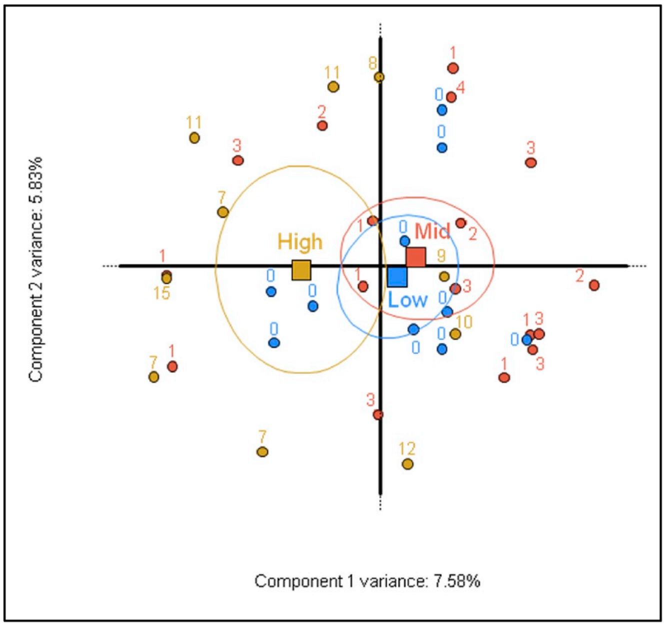


Fig. 10. R_V factor scores for Experiment 2: Constrained sorting task with natural stimuli, color-coded by musical experience (highly trained: orange; moderately trained: red; untrained: blue). Each dot represents a participant and the numbers corresponding to each dot represent the years of music training. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Component 3 (Fig. 8b) comes close to putting them in historical order, though Mozart and Beethoven overlap to a considerable degree. We think this dimension can be attributed to variability of texture, similar to Component 3 of Experiment 1 (Fig. 3b). Bach's pieces tend to stick with the same relatively closed texture for long periods of time, in contrast to those of Mozart and Beethoven, who often shift the texture in density and pitch range. We interpret Component 4 (Fig. 8b) as concerned with emotional engagement. Mozart (probably now more than in his own time) tends to be heard as elegant and above the fray, whereas Bach, and to an even greater extent Beethoven, tend to be heard as passionate and emotionally engaged. For Bach, this is especially true in performances by the pianists represented here (especially Barenboim, Arrau, and Richter), in contrast to a number of pianists who specialize in Bach, such as Glenn Gould and Rosalyn Tureck.

We now turn to the constrained sorts of Experiment 2 viewed in terms of the pianists (Fig. 9a & b). Keep in mind that the solution that underlies these figures is the same as the solution in Fig. 8; that is, all the individual points pertaining to excerpts are the same, but now the means ("barycenters") are calculated by grouping each pianist's points together, rather than each composer's. So, Component 1 (Fig. 9a) appears to indicate affinities between the pianists and particular composers: Barenboim with Beethoven, Arrau and Pirès with Bach, and Richter with Mozart. This last pairing is somewhat of a surprise, as in his career Richter was more typically associated with Beethoven and Bach than with Mozart. On the other hand, as Villemin (1999) noted, Richter was known for adapting his style to that of the composer he was playing, and so among the pianists here he may have been the best fit for Mozart. His Mozart in these excerpts was certainly among the most expressive performances of them. Component 2 (Fig. 9a) may concern overall heaviness of the texture, ranging from the relatively dark and heavy piano sound of the early Barenboim, to a moderately heavy sound of Arrau and Richter, to the very light sound of Pirès. This order parallels the progression from Bach through Beethoven to Mozart in Component 2 for composers (Fig. 8a). Component 3 (Fig. 9b) represents clarity of texture: Barenboim and Arrau (denser) versus Richter and Pirès (clear and lucid). (Component 3 for composers (Fig. 8b) contrasted the relatively dense Bach with the more open Mozart and Beethoven.) And we do not venture to interpret Component 4 (Fig. 9b).

One of the primary goals of this study was to verify the effectiveness of a sorting task and its analysis using DiSTATIS in musical style perception. The results showed that the sorting task could be successfully used to ascertain listeners' implicit knowledge of stylistic aspects, especially for untrained listeners. Many participants reported that they "had fun" sorting the excerpts, and that this task seemed less strenuous on them. Both music experts and non-experts performed similarly especially since the task did not rely on using technical vocabulary or any form of verbalization or quantification. On the other hand, knowing the basis of categorization of the melodies might help researchers understand the exact nature of cues (i.e., high-level or low-level) that each participant uses, and also to ascertain whether music training would influence the type of cues that listeners perceive. In a future study, researchers could ask participants to label each group of melodies after they complete both the free and constrained sorting tasks. Another reason for the untrained participants' competent performance on this task could be the use of excerpts from actual artistic performances, which contain a repository of cues pertaining to dynamics, texture, and so forth, not present in the MIDI versions of Experiment 1. Our study clearly showed that years of mere passive listening could facilitate the perception of such cues. One limitation of this study was that we did not assess the familiarity of our participants with each stimulus, and thus, we cannot estimate whether veridical knowledge (i.e., piece-specific information) helped the trained listeners in doing the task. Nevertheless, we did use relatively unfamiliar excerpts (see)Appendices A and B, and, most importantly, we did not see much differences in the relative performance of the untrained and the two trained groups. Also, all participants performed the free sort first followed by the constrained sort, as by definition it is impossible to counterbalance the order of presentation (see also Bigand, Filipic, et al., 2005; Bigand, Filipic, et al., 2005; Scott & Canter, 1997, using the same order). This meant that all participants doing the constrained sort were more familiar with the excerpts than in the free sort, thus potentially contributing to more coherent and converging responses in the second task. However, there is no confound here since all participants did the two tasks in the same sequence. Moreover, we re-randomized the order of the excerpts in the second task, so that participants had to re-categorize the excerpts based on the "new" constraints provided by the experimenter. A second limitation of the study is that we did not track the number of times participants heard each excerpt. For instance, Gingras et al. (2011), using the same sorting paradigm, found that the total number of times participants listened to each excerpt correlated significantly with their categorization accuracy. Also, an influence of musical expertise on problem-solving behavior in a musical puzzle task was reported by Tillmann, Bigand, and Madurell (1998); in particular, trained participants listened more often to the musical puzzle parts, but less often to the entire musical piece than did untrained participants. These overall findings convergently show that music experts tended to listen to the stimuli (or the parts of the puzzle individually) more often than the novices, which probably enhanced the experts' performance in the task. Building on these findings and our study, a future sorting experiment could investigate such a relationship between musical expertise and problem solving or perceptual strategies further. Future studies could also address if this task would be successful in discerning subtler and more nuanced aspects of musical style. For instance, would trained and untrained participants be able to sort melodies based on early versus late Beethoven's compositional style? Finally, an important follow-up experiment would be to investigate the effectiveness of this task when applied in a cross-cultural musical style perception study with expertise and familiarity as factors.

Appendix A

Experiment 1: MIDI Stimuli

No.	Composer	Key	Title
1	Bach	A	English Suite No. 1, Guigue 806
2		B^b	Partitas No. 1, BWV 825
3		C	Three-Part Invention, BWV 787
4		C-minor	French Suite No. 2, BWV 813
5		D	Prelude No. 5, BWV 850 (Well-tempered Piano I)
6		F	Little Fugue, BWV 556
7		G	French Suite No. 5, BWV 816
8	Mozart	A	Sonata K331, Allegro
9		B^b	Sonata K281, Allegro
10		C	Sonata K545, Allegro
11		C-minor	Sonata K457, Allegro assai
12		D	Sonata K576, Allegro
13		F	Sonata K280, Allegro
14		G	Sonata K283, Allegro
15	Beethoven	A	Sonata No. 2, Op. 2, Allegro
16		B^b	Sonata No. 11, Op. 22
17		C	Sonata in C, Op. 21, Allegro con brio
18		C-minor	Sonata No. 5, Op. 10 No. 1, Allegro
19		D	Sonata No. 7, Op. 10, Presto
20		F	Sonata No. 6, Op. 10 No. 2
21		G	Sonata No. 10, Op. 14, Allegro

Note. All are major keys except those explicitly designated as minor.

Appendix B

Experiment 2: Natural Stimuli

No.	Composer	Pianist	Key	Title
1	Bach	Arrau		Partita No. 2: Rondeaux Philips 434 904-2
2				Partita No. 3: Fantasia Philips 434 904-2
3				Partita No. 5: Praeambulum Philips 434 904-2
4		Barenboim		Goldberg Variations: Var. 18 Erato 741397T
5				Goldberg Variations: Var. 5 Erato 741397T
6				Goldberg Variations: Var. 6 Erato 741397
7		Pirès		Partita No. 1: Praeludium Philips 456 928-2
8				English Suite No. 3: Prelude Philips 456 928-2
9				French Suite No. 2: Allemande Philips 456 928-2
10		Richter		Das Wohltemperierte Clavier, Book I: Prelude 2 RCA Victor GD 60949
11				Das wohltemperierte Clavier, Book I: Prelude 5 RCA Victor GD 60949
12				Das wohltemperierte Clavier, Book II: Prelude 6 RCA Victor GD 60949
13	Mozart	Arrau		Sonata, KV 284: mvmt 1 Philips 432 306-2
14				Sonata, KV 330: mvmt 1 Philips 432 306-2
15				Sonata, KV 576: mvmt 1 Philips 432 306-2
16		Barenboim		Sonata, KV 281: mvmt 1 EMI CDZE 7 67294 2
17				Sonata, KV 533: mvmt 1 EMI CDZE 7 67294 2
18				Sonata, KV 311: mvmt 1 EMI CDZE 7 67294 2
19		Pirès		Sonata, KV 280: mvmt 1 DG 435 882-2
20				Sonata, KV 282: mvmt 3 DG 435 882-2
21				Sonata, KV 333: mvmt 1 DG 435 882-2

(*continued on next page*)

(*continued*)

No.	Composer	Pianist	Key	Title
22		Richter		Sonata, KV 283: mvmt 1 Philips 438 480-2
23				Sonata, KV 310: mvmt 1 Philips 422 583-2
24				Sonata, KV 457: mvmt 1 Philips 438 480-2
25	Beethoven	Arrau		Sonata No. 15, op. 28: mvmt 1 Philips 426 068-2
26				Sonata No. 21, op. 53: mvmt 1 Philips 426 068-2
27				Sonata No. 26, op. 81a: mvmt 3 Philips 426 068-2
28		Barenboim		Sonata No. 22, op. 54: mvmt 2 EMI 5 72912 2
29				Sonata No. 11, op. 22: mvmt 1 EMI 5 72912 2
30				Sonata No. 28, op. 101: mvmt 2 EMI 5 72912 2
31		Pirès		Sonata No. 14, op. 27, no. 2: mvmt 3 Erato 3984 27487 2
32				Sonata No. 17, op. 31, no. 2: mvmt 3 Erato 3984 27487 2
33				Sonata No. 23, op. 57: mvmt 3 Erato 3984 27487 2
34		Richter		Sonata No. 7, op. 10, no. 3: mvmt 1 Praga 354 022
35				Sonata No. 3, op. 2, no. 3: mvmt 1 Brilliant 92229/3
36				Sonata No. 31, op. 110: mvmt 2 Philips 454 949-2

References

Abdi, H. (2007). Distance. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 280–284). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781412952644.n142.

Abdi, H. (2007). Metric multidimensional scaling (MDS): Analyzing distance matrices. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 598–605). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781412952644.n279.

Abdi, H. (2010). Congruence: Congruence coefficient, R_V coefficient, and Mantel coefficient. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 222–229). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781412961288.n71.

Abdi, H., Valentin, D., Chollet, S., & Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DiSTATIS, theory and applications. *Food Quality and Preference, 18*(4), 627–640. https://doi.org/10.1016/j.foodqual.2006.09.003.

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433–459. https://doi.org/10.1002/wics.101.

Abdi, H., Williams, L. J., & Valentin, D. (2013). Multiple factor analysis: Principal component analysis for multi-table and multi-block data sets. *Wiley Interdisciplinary Reviews: Computational Statistics, 5*(2), 149–179. https://doi.org/10.1002/wics.1246.

Abdi, H., Williams, L. J., Valentin, D., & Bennani-Dosse, M. (2012). STATIS and DiSTATIS: Optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics, 4*(2), 124–167. https://doi.org/10.1002/wics.198.

Agres, K., Abdallah, S., & Pearce, M. (2018). Information-theoretic properties of auditory sequences dynamically influence expectation and memory. *Cognitive Science, 42*, 43–76. https://doi.org/10.1111/cogs.12477.

Atalay, B. V., & Placek, R. (1997). Machine versus human: Responding to the task of identifying eras for selected keyboard pieces. In A. Gabrielsson (Ed.), *Proceedings of the third triennial ESCOM conference* (pp. 521–526). Uppsala, Sweden: Uppsala University.

Beaton, D., Chin-Fatt, C. R., & Abdi, H. (2014). An ExPosition of multivariate analysis with the singular value decomposition in R. *Computational Statistics & Data Analysis, 72*, 176–189. https://doi.org/10.1016/j.csda.2013.11.006.

Beaton, D., Chin-Fatt, C. R., & Abdi, H. (2014). *Package " DistatisR ."*.

Bigand, E., Filipic, S., & Lalitte, P. (2005). The time course of emotional responses to music. *Annals of the New York Academy of Sciences, 1060*(1), 429–437. https://doi.org/10.1196/annals.1360.036.

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

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

Brown, R. (1981). Music and language. In *Documentary report of the Ann Arbor Symposium: National symposium on the applications of psychology to the teaching and learning of music*. Reston, VA: Music Educators National Conference.

Chin-Fatt, C. R., Beaton, D., & Abdi, H. (2013). *Package "MExPosition."*.

Chollet, S., Valentin, D., & Abdi, H. (2014). Free sorting task. In P. V. Tomasco, & G. Ares (Eds.), *Novel techniques in sensory characterization and consumer profiling* (pp. 207–227). Boca Raton, FL: Taylor & Francis.

Cope, D. (1991). *Computers and musical style*. Madison, WI: A-R Editions.

Crump, M. (2002). *A principal components approach to the perception of musical style* (Unpublished honors thesis). Canada: University of Lethbridge.

Dalla Bella, S., & Peretz, I. (2005). Differentiation of classical music requires little learning but rhythm. *Cognition, 96*, B65–B78. https://doi.org/10.1016/j.cognition.2004.12.005.

Dibben, N. (2001). What do we hear, when we hear music?: Music perception and music material. *Musicae Scientiae, 5*(2), 161–194. https://doi.org/10.1177/102986490100500203.

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science, 11*(3), 189–212. Retrieved from https://www.jstor.org/stable/2246110.

Dowling, W. J. (1986). Context effects on melody recognition: Scale-step versus interval representations. *Music Perception, 3*(3), 281–296. https://doi.org/10.2307/40285338.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology, 1*(1), 30–49. https://doi.org/10.1037/h0094275.

Eastlund, J. O. (1992). A multidimensional scaling analysis of musical style. *Journal of Research in Music Education, 40*(3), 204–215. https://doi.org/10.2307/3345682.

Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception, 18*(3), 275–296. https://doi.org/10.1525/mp.2001.18.3.275.

Gardner, H. (1973). Children's sensitivity to musical styles. *Merrill-Palmer Quarterly of Behavior and Development, 19*(1), 67–77. Retrieved from https://www.jstor.org/stable/23083794.

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., & McAdams, S. (2011). Perceiving musical individuality: Performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception, 40*, 1206–1220. https://doi.org/10.1068/p6891.

Gromko, J. E. (1993). Perceptual differences between expert and novice music listeners: A multidimensional scaling analysis. *Psychology of Music, 21*(1), 34–47. https://doi.org/10.1177/030573569302100103.

Grout, D. J., & Palisca, C. V. (1980). *A history of western music* (3rd ed.). New York, NY: W. W. Norton & Company, Inc.

Gundlach, R. H. (1932). A quantitative analysis of Indian music. *American Journal of Psychology, 44*(1), 133–145. https://doi.org/10.2307/1414960.

Hargreaves, D. J., & North, A. C. (1999). Developing concepts of musical style. *Musicae Scientiae, 3*(2), 193–216. https://doi.org/10.1177/102986499900300203.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology, 48*(2), 246–268. https://doi.org/10.2307/1415746.

Järvinen, T., Toiviainen, P., & Louhivuori, J. (1999). Classification and categorization of musical styles with statistical analysis and self-organizing maps. In A. Patrizio, G. A. Wiggins, & H. Pain (Eds.), *Proceedings of the AISB'99 symposium on musical*

creativity (pp. 54–57). Edinburgh, Scotland: The Society for the Study of Artificial Intelligence and Simulation of Behavior.
Krampe, R. T., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General, 125*, 331–359. https://doi.org/10.1037/0096-3445.125.4.331.
Krumhansl, C. L. (2015). Statistics, structure, and style in music. *Music Perception, 33*(1), 20–31. https://doi.org/10.1525/mp.2015.33.1.20.
Meyer, L. B. (1973). *Explaining music: Essays and explorations*. Berkeley, CA: University of California Press.
Meyer, L. B. (1989). *Style and music: Theory, history, and ideology*. Philadelphia, PA: University of Pennsylvania Press.
Miller, R. F. (1979). *An analysis of musical perception through multidimensional scaling* (Doctoral dissertation). Retrieved from ProQuest dissertations & theses global. (Order No. 8004230).
Morrison, S. J., & Yeh, C. S. (1999). Preference responses and use of written descriptors among music and nonmusic majors in the United States, Hong Kong, and the People's Republic of China. *Journal of Research in Music Education, 47*(1), 5–17. https://doi.org/10.2307/3345824.
Pascall, R. J. (1980). Style. In S. Stanley (Ed.), *The new grove dictionary of music and musicians* (pp. 316–321). London, UK: Macmillan.
R Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Raman, R., & Dowling, W. J. (2016). Real-time probing of modulations in South Indian classical (Carnātic) music by Indian and Western musicians. *Music Perception, 33*(3), 367–393. https://doi.org/10.1525/mp.2016.33.03.367.
Raman, R., & Dowling, W. J. (2017). Perception of modulations in South Indian classical (Carnātic) music by student and teacher musicians: A cross-cultural study. *Music Perception, 34*(4), 424–437. https://doi.org/10.1525/mp.2017.34.4.424.
Scott, M. J., & Canter, D. V. (1997). Picture or place? A multiple sorting study of landscape. *Journal of Environmental Psychology, 17*(4), 263–281. https://doi.org/10.1006/jevp.1997.0068.
Smaill, A., & Westhead, M. D. (1993). Automatic characterization of musical style. In M. Smith, A. Smaill, & G. A. Wiggins (Eds.), *Music education: An artificial intelligence approach* (pp. 157–170). Berlin, Germany: Springer-Verlag.
Storino, M., Dalmonte, R., & Baroni, M. (2007). An investigation on the perception of musical style. *Music Perception, 24*(5), 417–432. https://doi.org/10.1525/mp.2007.24.5.417.
Tekman, H. G., & Hortaçsu, N. (2002). Aspects of stylistic knowledge: What are different styles like and why do we listen to them? *Psychology of Music, 30*, 28–47. https://doi.org/10.1177/0305735602301005.
Thorisson, T. (1998). Comparison of novice listeners' similarity judgements and style categorisation of classic and romantic piano examplars. *Psychology of Music, 26*(2), 186–196. https://doi.org/10.1177/0305735698262006.
Tillmann, B., Abdi, H., & Dowling, W. J. (2004). Musical style perception by a linear auto-associator model and human listeners. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th international conference on music perception & cognition (ICMPC)* (pp. 189–191). Adelaide, Australia: Causal Productions.
Tillmann, B., Bigand, E., & Madurell, F. (1998). Local versus global processing of harmonic cadences in the solution of musical puzzles. *Psychological Research, 61*(3), 157–174. https://doi.org/10.1007/s004260050022.
Tyler, L. E. (1946). An exploratory study of discrimination of composer style. *The Journal of General Psychology, 34*(2), 153–163. https://doi.org/10.1080/00221309.1946.10544531.
Vignal, M. (1987). *Le dictionnaire de la musique*. Paris, France: Larousse.
Villemin, S. (1999). *Les grands pianistes*. Geneva, Switzerland: Georg.
Watt, R. J., & Ash, R. L. (1998). A psychological investigation of meaning in music. *Musicae Scientiae, 2*(1), 33–53. https://doi.org/10.1177/102986499800200103.
Wedin, L. (1969). Dimension analysis of the perception of musical style. *Scandinavian Journal of Psychology, 10*(1), 97–108. https://doi.org/10.1111/j.1467-9450.1969.tb00015.x.
Zivic, P. H. R., Shifres, F., & Cecci, G. A. (2013). Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences, 110*(24), 10034–10038. Retrieved from http://www.jstor.org/stable/42706125.