Principal Components Analysis of Musicality in Pitch Sequences

Richard Randall and Adam S. Greenberg

***Abstract*—**Musicality can be thought of as a property of sound that emerges when specific organizational parameters are present. We hypothesize that this property is not binary (where an auditory object *is* or *is not* a musical object), but rather exists on a continuum whereby some auditory objects may be considered more or less musical than other auditory objects. We suggest that identification of an auditory object as being more musical than another begins with a modularized analysis of features that coheres into a holistic interpretation. To explore this, we designed two experiments. In the first, 30 subjects evaluated 50 ten-tone sequences according to how musical they thought they were. A special stimulus set was designed that controlled for timbre, pitch content, pitch range, rhythm, note and sequence length, and loudness. Mean z-scored stimulus ratings showed significantly distinct groupings of musical versus non-musical sequences. In the second, a Principal Component Analysis (PCA) of the ratings yielded three components that explain a statistically significant proportion of variance in the ratings. The stimuli were analyzed in terms of parameters such as key correlation, range, and contour. These values were correlated with the eigenvalues of the significant PCA components in order to determine the dominant strategies listeners use to make decisions about musicality.

***Keywords*—**Musicality, principal component analysis, auditory perception.

I. Introduction

As a universal trait, humans are born with auditory predispositions that develop over time into musical knowledge and procedures. Adults with no formal musical training have developed the ability to make sophisticated judgments about music through years of exposure to a highly stable and organized auditory environment. Music cognition has been described as an interaction of bottom-up and top-down processes [1]. Auditory scene analysis constructs coherent objects from complex scenes, called auditory objects: the fundamental perceptual unit in audition [4][2][13]. When produced by a single source, auditory objects comprise multiple acoustic events that cohere to form a single auditory stream [27][28]. Stream formation is one of the key features of auditory scene analysis and is the mechanism by which we are able to attend to a specific speaker in a loud crowd (i.e. "the cocktail-party effect" [5]) or focus on a single instrument in a musical ensemble. Once a stream is formed, cognitive functions such as attention and memory are engaged to guide complex behaviors, allowing auditory objects to be processed by domain-specific processes of language or music [14][42]. This paper explores the idea that musicality is a property of auditory objects that emerges when certain organizational features are presented in specific ways. The perception of these musical objects requires multiple psychological processes to evaluate complex acoustic information. Moreover, this work address assumptions that musicality is a property that acoustic information either has or does not have.

Musical memory allows us to recall details about musical objects (e.g., recognition of a familiar melody), and contains schema that represent generalized information about our musical environment. In [1], Bharucha points out that humans have hierarchical schematic representations of musical features, such as metric and tonal organization. The temporal unfolding of a musical object allows its features to be matched with these schematic representations in an automatic way. Repeated exposure to a melody reinforces general mental representations: the stronger the representation, the easier melody recognition becomes [19]. Studies examining deficits in neurologically-impaired individuals observe that music perception is modular for two reasons. First, music is functionally distinct from language [35] and second, it relies on specialized modules that represent distinct music-cognitive processes [34]. Peretz and Coltheart [34] identified distinct neural networks associated with mid-level music processes such as meter, contour, interval, and tonal encoding. Only after an auditory object is processed by these modules can it engage what is referred to as the musical lexicon [34], which contains all of the information about music that one has been exposed to over one's lifetime. The recognition of familiar tunes, they suggest, is dependent on a selection procedure that takes place in the musical lexicon after an auditory object has been determined to be potentially musical by virtue of mid-level analysis of the aforementioned modular features. This idea was also explored in [40], which investigated the musical features that contributed to the recognition of familiar songs, supporting the idea that a song is a temporally unfolding perceptual object. Schulkind, et al. [40] found that familiar song recognition is a holistic process supported by multiple, interacting musical features. Their study was concerned with understanding what musical features correlate with the temporal position of object recognition. Our study is not concerned with familiar-object perception, but rather investigates subjects' perception of musicality in novel (never heard before) sequences.

We hypothesize that the perception of musicality is not binary (where an auditory object *is* or *is not* a musical object),

This research is supported by a Rothberg Research Award in Human Brain Imaging and by NIH grant T32-MH19983.

R. Randall is with the School of Music and the Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213 USA (412-268-4785; email: randall@cmu.edu).

A. S. Greenberg is with the Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI 53201 (email: agreenb@uwm.edu).

but rather exists on a continuum whereby some auditory objects may be considered more or less musical than other auditory objects. We suggest that identification of an auditory object as being more musical than another begins with a modularized analysis of features. The results of this analysis produce a percept that may be strongly or weakly associated with memories in the musical lexicon. Musicality, therefore, can be thought of as a property of sound that emerges when certain organizational features are presented in specific ways. By first probing an auditory object's degree of musicality and, second, investigating the features and processes that give rise to the degree of musicality, we can determine how listeners perceive an auditory object as musical.

Music theorists, empirical musicologists, and music psychologists have approached feature identification and analysis in a number of ways. One way is to examine structural features of existing music with the premise that, if a particular feature is significantly present in a specific way, then there is a mental representation or process that has developed in response to this feature. Another approach is to posit a general idea about "what music is" from the experimenter's experience. In both cases, features lead to testable hypotheses and the importance of the feature is established when observable behavior can be modulated by the manipulation of that feature. In both approaches, however, an auditory-object's status as "music" is taken *a priori*. While the assumption that Mozart's "Hunt Quartet" or an ethnic folk song is *axiomatically* music is appropriate in some contexts, it can inhibit us from understanding what a subject-oriented mental representation of music might look like. In the current study we seek to uncover an operational definition of music by first observing subjects' behavior and then examining the stimuli in an attempt understand how stimuli modulate behaviors. To test this, we asked subjects to listen to randomly generated pure-tone sequences and then rate each sequence on its musicality. Our first hypothesis is that subjects will be able to effectively group these sequences according to their musicality. Our second hypothesis is that a measure can be devised to quantitatively describe the psychological processes involved in judging the musicality of simple auditory stimuli. We created a profile for each sequence, comprised of a set of metrics that describe structural features commonly discussed in the music theory and music psychology literature. A principal component analysis (PCA) was performed on the ratings and the resulting components were correlated with the profiles to understand which features best account for the variance found in the musicality ratings. This approach allows for an exploration of the low-level auditory features that give rise to the perception of musicality in auditory objects. The following describes our two experiments and discussion is reserved for the end.

II. Experiment 1

A. Methods

1) Subjects

30 participants (17 female, 13 male) were recruited using the Carnegie Mellon University Center for Behavioral and Decision Research subject pool. Ages ranged from 19 to 58 years old with mean of 30.26 (*SD* 13). All subjects self-identified as having normal hearing and normal or corrected-to-normal vision. All were considered non-musicians with a mean 2.45 years of formal music training. All were native speakers of American English. No subjects reported having absolute pitch (AP) or knowledge of any family members with AP.

2) Stimuli

We used 50 sequences, each with 10 pure tones of 500ms duration, with an inter-stimulus-interval of 500ms. Each of the 10 tones was chosen randomly from the diatonic collection corresponding to the G-major scale (G4-F#5 or 392Hz-740Hz). Sequences were randomly transposed to all 12 pitch-class levels and then transposed back to the original G4-F#5 span creating an equal distribution of pitches across the chromatic scale.

3) Task

Subjects were asked to provide a rating of musicality for each sequence. They were instructed to rate each sequence on a Likert scale of 1 to 5. If they thought the sequence was very musical, they were asked to press the '5' key on the keyboard; if they thought the sequence was not musical at all, they were asked to press the '1' key on the keyboard. Subjects were asked to use the entire range of ratings and were given 1.5 seconds to respond. If they responded before the sequence had completed or failed to respond in 1.5 seconds, the response was not recorded and the experiment moved on to the next trial.

4) Design

Subjects completed a practice block of six trials after which they were asked if they understood the task. Subjects were given the option to complete additional practice trials or to move on to the main part of the experiment. The experiment was comprised of 15 blocks, each with 20 trials. 50 sequences were played eight times apiece for a total of 400 trials. Trials were presented in 20 blocks, with a forced 20-second break between each block. Stimuli were presented via headphones (Sennheiser HD210) at a fixed, comfortable listening volume (~82dBA-SPL). The paradigm ran on an Apple Mac Mini and responses were recorded on a standard computer keyboard. The paradigm was coded using MATLAB with Psychtoolbox extensions [3][33].

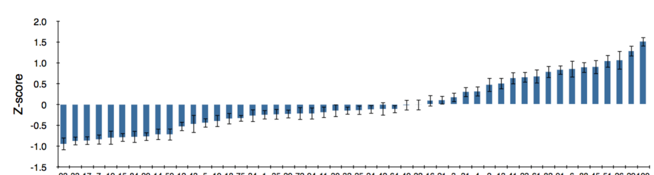

Fig. 1 Group (N=30) z-scored stimulus ratings of 50 sequences. Lower scores indicate less-musical sequences and higher scores indicate more-musical sequences.

5) Results

The results, shown in Fig. 1, show clear separability between the most musical and the least musical sequences, as well as a graded representation across the scale. Subjects used the entire scale in rating the sequences. For analysis, ratings were divided into two equal sized groups (lower 25 and upper

25). An ANOVA (alpha = 0.05) shows a significant difference between the two groups, $F(21,3)=17.45$, $p=0.019$. Fig. 2 shows the three most musical and the three least musical sequences for reference. This result demonstrates that the perception of musicality is inter-subjectively stable and is a quality that varies across stimuli.

Fig. 2 The three-most (red) and three-least (blue) musical sequences with z-scores.

III. Experiment 2

A. Methods

We identify five feature categories that figure predominantly in the music-theoretic and music-psychological literature: interval, contour, tonality, motive, and entropy. This list is not complete, nor comprehensive, but it is representative. Each feature is represented as a metric so we can understand the consequences of its variation. Metrics were derived from each sequence using *Music21* [7], *Midi Toolbox* [10], standard statistical analysis software (MATLAB and SPSS), and custom scripts. Like *Humdrum Toolkit* [15] before it, *Music21* and *Midi Toolbox* come with a set of analytic tools or routines grounded in recent psychological and music-theoretic literature that effectively explore certain musical features. Other features we explored required more creative treatments for them to be useful in our analysis. Below are descriptions of the features we used, their psychological and music theoretical context, and some general predictions about how they might relate to the results of experiment 1.

1) Intervals

A pitch interval is the distance in semitones between two tones and a melodic interval is the distance between two adjacent pitches in a sequence. We created an intervallic profile for each sequence by calculating the mean interval size (the average of all consecutive melodic intervals), standard deviation (SD) of the mean, and range [19]. In addition, we calculated the melodic interval variability (MIV), a version of the coefficient of intervallic variation of a sequence used in music/language research [32].

2) Contour

Contour refers to the shape of musical materials (e.g., pitches, rhythms, timbres, tempi,) [29]. In a simple melodic context where variation occurs only in the pitch domain, contour refers to shape of the sequence as the pitch ascends and/or descends in frequency space. Analysis of the Essen Folk Song collection reveals that the predominant contour for 10-note sequences is a clear arch shape that gradually rises from a starting point, peaks midway, and returns to end at the original or near-original starting point [16]. Using [16]'s 10-note arch as an archetype, we defined a function that returns a value that represents how well each sequence fits this archetype (e.g. correlation). The contour correlation value ranges from 1 to 0, where 1 represents equivalence between a sequence and the archetype. The contour of a sequence is usually easier to remember than exact interval information [9][8]. Given the strength of Huron's findings, we predict that sequences that more closely fit this archetype will be perceived as being more musical than sequences that do not fit the archetype.

3) Tonality

Tonality is the property whereby tones in a scale are hierarchically organized around a central pitch. Key finding algorithms correlate a musical excerpt with a key according to a probabilistic distribution based on the Krumhansl tonal profile [22]. Supporting [34]'s findings, recent behavioral research asserts that identification of a tonal center is an elemental process at the core of how all listeners experience music [11]. We used the key finding algorithms implemented in *Music21* to find key centers of each sequence [7]. While the algorithms are in many ways similar, each is methodologically different and the methods described by [7] are not comprehensive with respect to all available approaches. Since the sequences we used are only 10 notes in length and the key-finding algorithms are based on statistics whose power increases with the number of pitches, we selected only the highest correlation for each sequence. The actual key correlation of a sequence is irrelevant for this study; our *maximum key correlation* metric represents the likelihood that a certain key, or tonal center, can be inferred. Since [34] proposes that the inference of key is necessary for an auditory object to be musical, we hypothesize that sequences with higher key correlations will be considered more musical than those with lower correlations.

4) Motive

The use of motives (small, related, and easily remembered musical ideas) is an important component of the overall musical experience, as motives allow listeners to conflate smaller musical ideas into a single larger concept. Following [4], when sounds cohere they form an auditory stream that is separable from other sounds or other streams. However, [49] maintains that coherence allows multiple streams to be compared with one another, thus promoting a higher-level of perceptual organization – a property [49] shows is commonly found in music of the Western classical canon. In this sense, motivic coherence is similar to compression, whereby complex auditory scenes comprising multiple auditory objects are made easier to interpret and remember by virtue of a similarity between objects. While there is a literature addressing the segmentation of music into smaller units, such approaches rely on features (e.g., such as varied rhythm) that our melodic stimuli do not possess. Because our sequences are relatively feature-poor, we used prediction by partial matching (PPM) to measure the compressibility of each sequence. PPM is an adaptive statistical compression technique that predicts the *n*th symbol in a string by using a context that optimally varies in length according to its predictive success [26]. We used a PPM encoder on consecutive intervals in each sequence and limited the context size to a maximum of 5. Since motive-rich music is more highly valued than motive-poor music by some music scholars [31][6][12][49], we expect that greater compressibility will correlate with greater musicality.

5) Entropy

Probabilistic entropy has a rich history in music scholarship largely centered on style analysis [48][20][41]. It has also been used to model melodic complexity, musical expectation, and aesthetic experience [25][10][23]. Entropy is closely related to ideas of compression found in our PPM analysis and the probabilistic distribution found in our key finding analysis. Relative entropy (Hr) is a convenient and standard metric for characterizing how varied the distribution of symbols (pitches) is given a fixed alphabet (the diatonic scale). It is difficult to make predictions about how varied levels of uncertainty might affect musicality. We might predict that a 10-note sequence composed using a single pitch might be equally as musical as a 10-note sequence employing all seven diatonic pitches. Nevertheless, because of its continued appearance in music-related studies, we included relative entropy in the sequence profile.

Fig. 3 shows the aforementioned metrics associated with the three most- and least-musical sequences.

D93	D33	D17		D26	D29	D100
-0.95	-0.88	-0.87	Z-Score	1.06	1.28	1.51
9	10	10	Range	4	10	8
3.34	5.56	4.44	Mean	1.33	2.36	1.76
2.56	3.43	3.06	SD	0.94	3.33	2.33
1.31	0.62	0.62	MIV	0.71	0.71	0.76
0.71	0.82	0.56	Key	0.64	0.82	0.71
0.67	0.82	0.71	Contour	0.57	0.71	0.49
0.65	0.86	0.78	Hr	0.48	0.81	0.97
43.81	64.48	53.97	PPM	36.98	54.53	51.68

Least Musical ← → Most Musical

Fig. 3 Feature metrics for the three-most and three-least musical sequences shown.

IV. Results

The first three intervallic features (Range, Mean, and SD) are strongly negatively correlated ($p=<0.001$) with z-mean scores. Subjects found that sequences with smaller range, smaller mean-interval size, and smaller standard deviation of the mean to be more musical (Table 1). This confirms our prediction and supports the Huron's detailed work showing that cross-culturally, sequences privilege small intervals [19]. Interestingly, features such as MIV, Contour, Key, Hr, and PPM do not significantly correlate with the ratings.

We hypothesize that subjects' ratings will have uncorrelated principal components that will each correlate with the different musical features described above. That is, components are separable in terms of these features suggesting that these features are used, to varying degrees, in combination when subjects make musicality judgments of pitch sequences. PCA of the ratings was performed and permutation testing showed that only the first three components were significant, explaining a combined 38% of the variance; however, this is a conservative measure. Because information is often embedded in additional components, we include the top five. The feature profile of each tune was correlated with the eigenvalues of the significant PCA components in order to identify dominant strategies listeners use to make decisions about musicality. The PCA analysis is shown in Fig. 4.

Table 1
Significant correlations between features and ratings.

		Zscore
Zscore	Pearson Correlation	1
	Sig. (2-tailed)	
	N	50
Mean	Pearson Correlation	-.619**
	Sig. (2-tailed)	.000
	N	50
Range	Pearson Correlation	-.491**
	Sig. (2-tailed)	.000
	N	50
SD	Pearson Correlation	-.697**
	Sig. (2-tailed)	.000
	N	50

Table 2 shows correlations between eigenvalues of the first five components and the eight metrics. Significant correlations are highlighted in red. Of the eight, only intervallic features of Range, Mean, and SD correlate significantly with the first component. Mean also correlates strongly with the second component as does Key. Hr and PPM appear less strongly followed by MIV. Contour correlates with the third component followed closely by a reappearance of PPM.

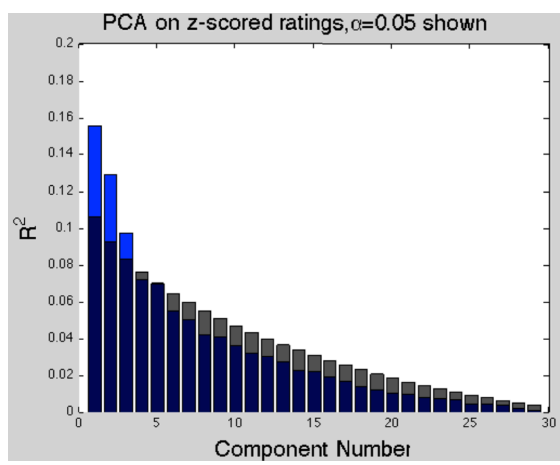

Fig. 4 A principal components analysis (PCA) on subjects' ratings returns three significant uncorrelated components explaining 38% of the variance. Gray bars indicate critical R^2 values from permutation testing; blue bars indicate R^2 values from PCA.

V. Discussion

The goal of this study was to understand the conditions under which randomly generated sequences are organized into (or perceived as) auditory objects. Despite the extensive work that has been done on auditory stream formation, few studies have examined how this process plays out in novel musical contexts. Notable treatments address how bistable auditory streams are formed in polyphonic music and how acoustic features, such as timbre, contribute to object formation [47][24][45]. Unlike previous studies that present multi-stable auditory scenes, this study presents a single stream in a non-competitive auditory environment. The streams were controlled for range, timbre, tempo, and event duration, so as to minimize the possibility that any one stream could be interpreted as multi-stable. However, the perception of multiple streams and how it contributes to musical object formation remains an open question.

TABLE II
SIGNIFICANT CORRELATIONS (IN RED) BETWEEN FEATURES AND COEFFICIENTS

		Comp1	Comp2	Comp3	Comp4	Comp5
Mean	Pearson Correlation	.647**	.445**	-.028	-.076	.001
	Sig. (2-tailed)	.000	.001	.844	.598	.995
	N	50	50	50	50	50
Range	Pearson Correlation	.576**	.170	.233	.106	-.048
	Sig. (2-tailed)	.000	.237	.104	.463	.741
	N	50	50	50	50	50
SD	Pearson Correlation	.652**	.170	-.009	.230	.205
	Sig. (2-tailed)	.000	.239	.952	.108	.152
	N	50	50	50	50	50
Key	Pearson Correlation	.033	.445**	.039	-.168	.026
	Sig. (2-tailed)	.822	.001	.787	.244	.860
	N	50	50	50	50	50
Hr	Pearson Correlation	.246	.412**	.262	.012	-.190
	Sig. (2-tailed)	.085	.003	.066	.936	.186
	N	50	50	50	50	50
PPM	Pearson Correlation	.093	.358*	.451**	.065	-.267
	Sig. (2-tailed)	.521	.011	.001	.652	.061
	N	50	50	50	50	50
MIV	Pearson Correlation	-.042	-.327*	-.005	.286*	.293*
	Sig. (2-tailed)	.771	.021	.973	.044	.039
	N	50	50	50	50	50
Contour	Pearson Correlation	.060	-.244	.494**	.033	-.134
	Sig. (2-tailed)	.677	.087	.000	.820	.354
	N	50	50	50	50	50

Additional analysis was performed in an attempt to understand how select musical features interact with each other to support the perception of musicality. Fig. 5 shows the linear combination of metrics that significantly correlate with component 1 plotted against the eigenvalues. Musicality ratings are color coded with red as most musical and blue as least musical. A clear pattern is observed whereby higher musicality ratings cluster towards smaller values and lower musicality ratings cluster upward toward larger values. This is shown more clearly in Fig. 6, which limits the linear combination to significantly correlated metrics of the ten most and least musical.

Comparing Figs. 5 and 6 we see that there is an organizational difference between high and low musical sequences. For the significant metrics, their strong bilateral separation between high and low musical sequences along the *x* axis shows that subjects are using these features (Mean, Range, and SD) consistently to make musicality judgments. Furthermore, separable groups along the *y* axis show that as composite values of the metrics increase, so too does variance.

In contrast, Fig. 7 plots the linear combination of metrics that *do not* significantly correlate with component 1 against its eigenvalues. The pattern is more diffuse with a slight bilateral distribution of high versus low musicality across the *x* axis, but no grouping along the *y* axis. We interpret this as evidence that the non-significant metrics (Key, Hr, PPM, MIV, and Contour) are not contributing meaningfully to the variance at this level.

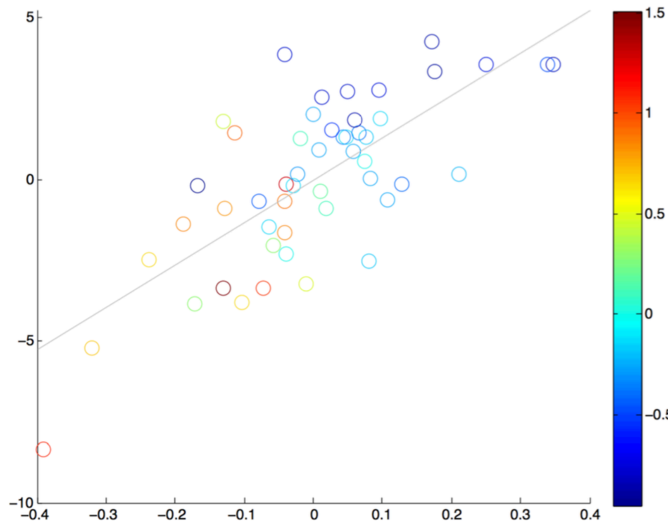

Fig. 5 Linear combination of metrics significantly correlated with component 1 plotted against eigenvalues. Z-score for each sequence is color-coded (red = high, blue=low).

Our analysis shows that common features used to describe musical works can explain how subjects determine the degree to which an auditory object is musical in a limited way. It raises interesting questions about the ontological role of these (and other) analytic descriptions. For instance, we might claim that while key membership is important for determining whether or not a sequence of tones is a *musical* sequence of tones, that quality might only be important if specific intervallic features are present as well. We might also assert that while contours in music exhibit regular patterns (e.g., the "melodic arch"), such patterns may not play a particularly strong role in musical auditory object formation, or the role they play might be dependant on the presence other features.

PCA has been used by [39] and [37] in music perception studies to simplify high-dimensional models. Here, we use PCA in an exploratory way to identify components that explain the variance found in the data. All features described in section III-A significantly correlate with one or more of the components. However, components, even if significant, may not be statistically independent from one another; therefore, we cannot make any claims about the independence of components: nor can we make claims about ordering. In other words, the fact that component 1 is most strongly correlated with interval analysis does not mean that it does not impact component 2.

Music perception in its most basic sense is dependent on listeners' ability to recognize the presence or absence of structural features and to build a representation of auditory objects. This study explores how such features relate to the perceived musicality of a pitch sequence. It combines perspectives from music theory, computational musicology, and music perception to identify a set of features that, while not exhaustive, all represent musical experiences. Experiment 2 is constrained by the set of features we have chosen. Whether or not a feature has explanatory power, of course, depends on the encoding of that feature. It is possible that we are limited in thinking about musical features in specific ways, finding just enough evidence to continue considering them an important part of our mental representations of music.

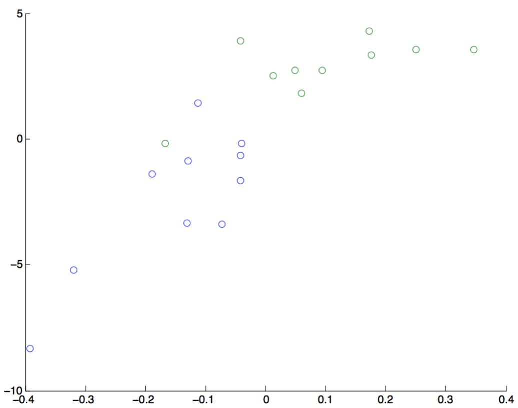

Fig. 6 Linear combination of metrics significantly correlated with component 1 plotted against eigenvalues for the 10 most musical (blue) and 10 least musical (green) sequences.

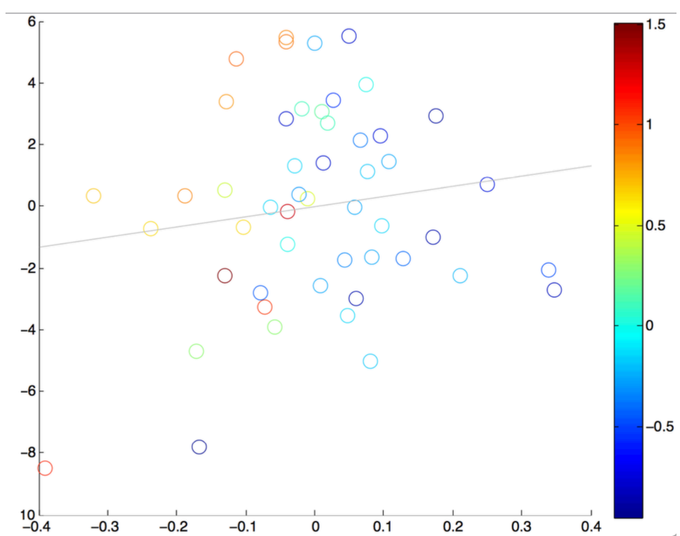

Fig. 7 Linear combination of metrics not significantly correlated with component 1 plotted against eigenvalues. Z-score for each sequence is color coded (red = high, blue=low).

Our approach to explicate these mental representations provides a new way to examine music. This study sought to understand the boundary conditions for how auditory objects can become musical objects. We show that musicality is a variable quality that auditory objects possess in greater or lesser degrees. We also show that there is considerable intersubjective agreement about what constitutes a highly musical object. We see that musicality is a property of auditory objects that emerges under specific conditions, and exists on a continuum. Placement on this continuum is the result of multiple interacting features. Our results are informative for future music studies and our empirically derived collection of 10-tone sequences is fertile territory for exploration. Future work includes identifying the neural correlates of the varied perception of auditory objects.

References

[1] J. J. Bharucha, "Music cognition and perceptual facilitation: A connectionist framework," *Music Perception*, pp. 1–30, 1987.

[2] J. K. Bizley and Y. E. Cohen, "The what, where and how of auditory-object perception," *Nature Reviews Neuroscience*, vol. 14, no. 10, pp. 693–707, 2013.

[3] D. H. Brainard, "The psychophysics toolbox," *Spatial vision*, vol. 10, pp. 433–436, 1997.

[4] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[5] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the acoustical society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[6] E. T. Cone, "On derivation: Syntax and rhetoric," *Music Analysis*, vol. 6, no. 3, pp. 237–255, 1987.

[7] M. S. Cuthbert and C. Ariza, "Music21: A toolkit for computer-aided musicology and symbolic music data," 2010.

[8] W. J. Dowling and D. S. Fujitani, "Contour, interval, and pitch recognition in memory for melodies," *The Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 524–531, 1971.

[9] W. J. Dowling, "Scale and contour: Two components of a theory of memory for melodies.," *Psychological review*, vol. 85, no. 4, p. 341, 1978.

[10] T. Eerola and P. Toiviainen, "MIDI toolbox: MATLAB tools for music research," 2004.

[11] M. M. Farbood, G. Marcus, and D. Poeppel, "Temporal dynamics and the identification of musical key.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 39, no. 4, p. 911, 2013.

[12] R. O. Gjerdingen, "Revisiting Meyer's' grammatical simplicity and relational richness,'" 1992.

[13] T. D. Griffiths and J. D. Warren, "What is an auditory object?," *Nature Reviews Neuroscience*, vol. 5, no. 11, pp. 887–892, 2004.

[14] L. L. Holt and A. J. Lotto, "Speech Perception Within an Auditory Cognitive Science Framework," *Current Directions in Psychological Science*, vol. 17, no. 1, pp. 42–46, Feb. 2008.

[15] D. Huron, "The Humdrum Toolkit: Software for Music Research," *Center for Computer Assisted Research in the Humanities, Ohio State University, copyright*, vol. 1999, 1993.

[16] D. Huron, "The melodic arch in Western folksongs," *Computing in Musicology*, vol. 10, pp. 3–23, 1996.

[17] D. Huron, "Music research using Humdrum: A user's guide," *Center for Computer Assisted Research in the Humanities, Stanford, California*, 1999.

[18] D. Huron, "Music information processing using the Humdrum Toolkit: Concepts, examples, and lessons," *Computer Music Journal*, vol. 26, no. 2, pp. 11–26, 2002.

[19] D. B. Huron, *Sweet anticipation: Music and the psychology of expectation*. MIT press, 2006.

[20] L. Knopoff and W. Hutchinson, "Entropy as a measure of style: The influence of sample length," *Journal of Music Theory*, vol. 27, no. 1, pp. 75–97, 1983.

[21] C. L. Krumhansl and M. A. Schmuckler, "Key-finding in music: An algorithm based on pattern matching to tonal hierarchies," in *Mathematical Psychology Meeting*, 1986.

[22] C. L. Krumhansl, "Perceptual structures for tonal music," *Music Perception*, pp. 28–62, 1983.

[23] E. H. Margulis and A. P. Beatty, "Musical style, psychoaesthetics, and prospects for entropy as an analytic tool," *Computer Music Journal*, vol. 32, no. 4, pp. 64–78, 2008.

[24] S. McAdams, "Perspectives on the contribution of timbre to musical structure," *Computer Music Journal*, vol. 23, no. 3, pp. 85–102, 1999.

[25] L. B. Meyer, *Emotion and meaning in music*. University of chicago Press, 2008.

[26] A. Moffat, "Implementing the PPM data compression scheme," *Communications, IEEE Transactions on*, vol. 38, no. 11, pp. 1917–1921, 1990.

[27] B. C. Moore and H. Gockel, "Factors influencing sequential stream

segregation," *Acta Acustica United with Acustica*, vol. 88, no. 3, pp. 320–333, 2002.

[28] B. C. Moore and H. E. Gockel, "Properties of auditory stream formation," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1591, pp. 919–931, 2012.

[29] R. Morris, "Composition with Pitch-classes," *Yale University*, 1987.

[30] R. D. Morris, "New directions in the theory and analysis of musical contour," *Music Theory Spectrum*, vol. 15, no. 2, pp. 205–228, 1993.

[31] E. Narmour, *The analysis and cognition of melodic complexity: The implication-realization model*. University of Chicago Press, 1992.

[32] A. D. Patel, *Music, language, and the brain*. Oxford university press, 2010.

[33] D. G. Pelli, "The VideoToolbox software for visual psychophysics: Transforming numbers into movies," *Spatial vision*, vol. 10, no. 4, pp. 437–442, 1997.

[34] I. Peretz and M. Coltheart, "Modularity of music processing," *Nature neuroscience*, vol. 6, no. 7, pp. 688–691, 2003.

[35] I. Peretz and K. L. Hyde, "What is specific to music processing? Insights from congenital amusia," *Trends in cognitive sciences*, vol. 7, no. 8, pp. 362–367, 2003.

[36] I. Peretz, "The nature of music from a biological perspective," *Cognition*, vol. 100, no. 1, pp. 1–32, 2006.

[37] L. Quinto, W. F. Thompson, and A. Taylor, "The contributions of compositional structure and performance expression to the communication of emotion in music," *Psychology of Music*, p. 0305735613482023, 2013.

[38] H. Schaffrath and D. Huron, "The Essen folksong collection in the humdrum kern format," *Menlo Park, CA: Center for Computer Assisted Research in the Humanities*, 1995.

[39] E. G. Schellenberg, "Simplifying the implication-realization model of melodic expectancy," *Music Perception: An Interdisciplinary Journal*, vol. 14, no. 3, pp. 295–318, 1997.

[40] M. D. Schulkind, R. J. Posner, and D. C. Rubin, "Musical Features That Facilitate Melody Identification: How Do You Know It's '"Your"' Song When They Finally Play It?," *Music Perception: An Interdisciplinary Journal*, vol. 21, no. 2, pp. 217–249, 2003.

[41] J. L. Snyder, "Entropy as a measure of musical style: the influence of a priori assumptions," *Music Theory Spectrum*, vol. 12, no. 1, pp. 121–160, 1990.

[42] D. L. Strait, N. Kraus, A. Parbery-Clark, and R. Ashley, "Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance," *Hearing research*, vol. 261, no. 1, pp. 22–29, 2010.

[43] D. Temperley, "What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered," *Music Perception*, pp. 65–100, 1999.

[44] D. Temperley, *Music and probability*. MIT Press Cambridge^ eMass Mass, 2007.

[45] M. Uhlig, M. T. Fairhurst, and P. E. Keller, "The importance of integration and top-down salience when listening to complex multi-part musical stimuli," *Neuroimage*, vol. 77, pp. 52–61, 2013.

[46] N. L. Wood and N. Cowan, "The cocktail party phenomenon revisited: attention and memory in the classic selective listening procedure of Cherry (1953).," *Journal of Experimental Psychology: General*, vol. 124, no. 3, p. 243, 1995.

[47] J. K. Wright and A. S. Bregman, "Auditory stream segregation and the control of dissonance in polyphonic music," *Contemporary Music Review*, vol. 2, no. 1, pp. 63–92, 1987.

[48] J. E. Youngblood, "Style as information," *Journal of Music Theory*, vol. 2, no. 1, pp. 24–35, 1958.

[49] L. M. Zbikowski, "Musical coherence, motive, and categorization," *Music Perception: An Interdisciplinary Journal*, vol. 17, no. 1, pp. 5–42, 1999.