Perceptual basis of evolving Western musical styles

Pablo H. Rodriguez Zivic[a,1], Favio Shifres[b], and Guillermo A. Cecchi[c]

[a]Computer Science Department, University of Buenos Aires, 1428 Buenos Aires, Argentina; [b]Laboratory for Music Experience Study, Faculty of Fine Arts, National University of La Plata, 1900 La Plata, Argentina; and [c]Computational Biology Center, T. J. Watson IBM Research Center, Yorktown Heights, NY 10598

The brain processes temporal statistics to predict future events and to categorize perceptual objects. These statistics, called expectancies, are found in music perception, and they span a variety of different features and time scales. Specifically, there is evidence that music perception involves strong expectancies regarding the distribution of a melodic interval, namely, the distance between two consecutive notes within the context of another. The recent availability of a large Western music dataset, consisting of the historical record condensed as melodic interval counts, has opened new possibilities for data-driven analysis of musical perception. In this context, we present an analytical approach that, based on cognitive theories of music expectation and machine learning techniques, recovers a set of factors that accurately identifies historical trends and stylistic transitions between the Baroque, Classical, Romantic, and Post-Romantic periods. We also offer a plausible musicological and cognitive interpretation of these factors, allowing us to propose them as data-driven principles of melodic expectation.

pattern recognition | psychology | computational cognition | culturomics

Like any cultural phenomenon, musical style evolves through history in recognizable form, driven by technical and societal changes, in part, but largely following its own dynamics. An essential question for both music theory and cognitive science is to what extent, and how, these putative dynamics are constrained by fundamental perceptual structures. Style can be recognized by characteristic uses of form, texture, harmony, melody, and rhythm (1), yielding patterns that can then be mined to reveal hidden structures in music (2). Melody, in particular, defined as pitched sounds arranged over time in accordance with given cultural conventions and constraints, is an essential component of style (1). When played in a sequence, notes form perceptual relations with each other. Specifically, two adjacent notes define a melodic interval, measured as the distance in semitones between the second note and the first note. Current cognitive theories, such as implication–realization (IR) theory (3), and experimental results on short-term music expectation (4–6) propose that only three consecutive notes—two melodic intervals, or one bigram—are required to induce strong expectations on melodic continuations. This connection between a fundamental element of style and basic perceptual mechanisms provides a departing point to understand how style dynamics may be cognitively constrained.

In this context, our contributions are threefold. We first take advantage of the availability of a large music dataset (7) and propose a unique method for analyzing bigram probability distributions that exploits its latent structure to identify accurately the transitions between the Baroque, Classical, Romantic, and Post-Romantic periods. Second, we propose a cognitive interpretation of results, consistent with current musicological theories. Finally, we show that the prediction power of our model greatly exceeds that of the IR theory, letting us conclude that this method yields data-driven principles that can extend current cognitive theories on short-term music expectation.

Results

We selected the Peachnote corpus (7) as a source of data for our analysis. It consists of the number of times each melodic interval

pattern was used over each year between 1730 and 1930. Using this information, we first analyze each year's conditional probability distribution of melodic intervals $P_y(i_2|i_1)$ (*SI Materials and Methods*). Because the data span a period during which music underwent substantial changes, we expect to observe fluctuations on this probability distribution reflecting musical evolution. Thus, we reasoned that styles should correspond to prototypical distributions that can be identified by an appropriate clustering algorithm; in our case, we decided to use k-means (ref. 8, p. 202). To measure the distance between probability distributions, we used the Frobenious norm of its matrix representation. With respect to the number of clusters, we tested several values of k, finding an optimum value at $k = 5$.

Fig. 1 shows a visualization of the clustering analysis, computed as follows: After having computed the k-means clustering, each data point was sorted by its corresponding year, which was not used in the clustering procedure. Then, for each cluster index c, we plotted a function of time, which measures the proportion of matrices clustered in the same group within a temporal window. Formally,

$$f_c(y) = \frac{1}{\tau} \sum_{i=y-\tau+1}^{y} \delta(CI(y) - c),$$

where $CI(y)$ represents the cluster index assigned by the algorithm to the year y, τ is the size in years of the smoothing window, and δ is Kronecker's delta function.* In this case, we used a 10-y lag.

The clusters yielded by the conditional distribution group neighboring years consistently by assigning them the same cluster index. As a reference, we plotted three dotted lines on the years 1750, 1830, and 1900. These years are of major importance to music history because they correspond to the approximate transition between different musical styles. The Baroque period ended between approximately 1750 and 1770 (1), the Classical period spanned from the end of the Baroque to ~1830, and the Romantic period ended around 1900 (9).

The results in Fig. 1 provide evidence that the conditional probability distribution of melodic intervals contains useful information for style recognition. However, the clustering analysis offers limited insights into the underlying mechanisms that drive style evolution. To address this, we present a factor-based dimensionality reduction technique, and we show that the

Fig. 1. Visualization of the clusters for the conditional distribution of melodic intervals. Each shaded area corresponds to a different cluster, and its corresponding line represents the proportion of years assigned to it within a 10-y smoothing window. Vertical dashed lines correspond to the approximate boundaries between Baroque, Classical, and Romantic periods. A fifth cluster was removed because it was a noise cluster with only three elements.

results can be interpreted in terms of musicological and cognitive theories.

We show in Fig. 1 that well-defined temporal trends are captured by the bigram statistics. A natural step after the recognition of clusters is to identify understandable patterns that explain them. Given that probability distributions are nonnegative by definition, it is convenient for the purpose of interpretation to use a dimensionality reduction approach that preserves this feature in the factors that represent the data; for this, we chose nonnegative matrix factorization (NMF) (10).

Recent work on NMF demonstrates relationships with certain families of *k*-means clustering (11, 12). Specifically, the simple *k*-means algorithm was proven equivalent to orthogonal NMF (13) in terms of the yielded clustering. Both NMF versions find a decomposition minimizing a cost function; however, orthogonal NMF penalizes correlated solutions. This is a twofold advantage: On the one hand, orthogonal NMF is a natural extension of the analysis presented in the previous section because it builds a decomposition that yields the clustering shown in Fig. 1; on the other hand, the orthogonality restriction aids in the interpretation of the decomposition.

Preliminary analysis showed that even though the conditional distribution of melodic intervals yields a better clustering, the joint distribution, $P(i_1, i_2)$, is much more interpretable. This led us to build an equivalent version of joint and conditional decompositions (*SI Materials and Methods*). We coded the corpus into a matrix T, whose rows correspond to a flattened coding of each year's data. We then ran orthogonal NMF on T (Fig. 2), yielding matrices W and H, such that $T \approx WH$.

In this context, the decomposition's interpretation is straightforward: W contains a flattened coding of the factors that explain the distribution, and H is their projection over time. Intuitively, W corresponds to a set of relevant characteristics of the styles found using *k*-means, and W is their weight over time. Because both W and H are nonnegative and orthogonalized, the projection over time of each factor can be thought of as the degree of its participation in that year's distribution.

Discussion

In the previous section, we showed that we can detect stylistic trends with the distribution of melodic interval bigrams. Moreover,

we also found that some aspects of the distribution are more prominent in the different periods. The next step is to ask whether we can interpret the constructed model or if it is just a tool to predict style epochs. From a musicological point of view, we will interpret our results by analyzing the factors in terms of the musical structure they represent and match them with characteristics of the corresponding style. From a cognitive perspective, we will relate the empirical probability distributions to the expectancies of the listeners.

Musicological Interpretation. We start with the Baroque period. The first factor of Fig. 2 appears to be dominated by intervals of two semitones, but there are also intervals of one semitone. We hypothesize that this factor could be due to adjacent diatonic movement on a scale, which is a characteristic of late Baroque music (ref. 14, pp. 220–222). Taking advantage of the fact that a major scale is built on an established intervallic pattern,[†] we can use each of its bigrams to create a binary matrix. This matrix corresponds to all possible ways of playing three adjacent notes of a scale, also called diatonic notes (Fig. 3). There is a remarkable similarity between this matrix and the structure of the joint version of the Baroque factor. Note that even though stepwise movement is ubiquitous in tonal music, it is attributed to the Baroque period. This behavior is achieved by the orthogonalization constraint imposed on the factors, making each bigram be attached only to the cluster in which it appears the most.

The Classical factor is a red dot representing the double unison, namely, a melodic interval whose size is equal to zero. There is also a slight weight over any interval followed by a unison and over a unison followed by any interval. This is consistent with the Classical preference for unison both found in the study by Simonton (15) and controlled with another corpus (*SI Materials and Methods*).

The factor associated with the Romantic period reflects a wider intervallic use and the avoidance of scale adjacent notes. Its structure is shared and extended by the Post-Romantic factor, which uses a wider intervallic repertoire. Moreover, in both factors, the most likely intervals are between three and five semitones

---

[†]The intervallic pattern that corresponds to a major scale is 2212221, and that corresponding to a minor scale is 2122122. Note that it is the same shifted pattern.

in size, which are the building blocks of more complex structures in musical language, such as chords. Finally, this structure extended by the Post-Romantic factor can be understood as an exploration of new sounds, a phenomenon that was coined "emancipation of dissonance" by Schoenberg (16).

Cognitive Interpretation. We now discuss our findings in the context of perceptual and cognitive concepts, based on the idea of melodic expectation. Expectation is considered to be a central feature in melodic understanding (17, 18), spanning a variety of time scales (19). Specifically, there is experimental evidence of strong expectancies over the scope of melodic interval bigrams (4–6, 20, 21), similar to what is observed with other perceptual modalities, such as the law of good continuation in visual Gestalt (22).

The present analysis is built on the idea that expectation and probabilities distribution must be related; otherwise, listeners would constantly expect unlikely events. Therefore, variation of
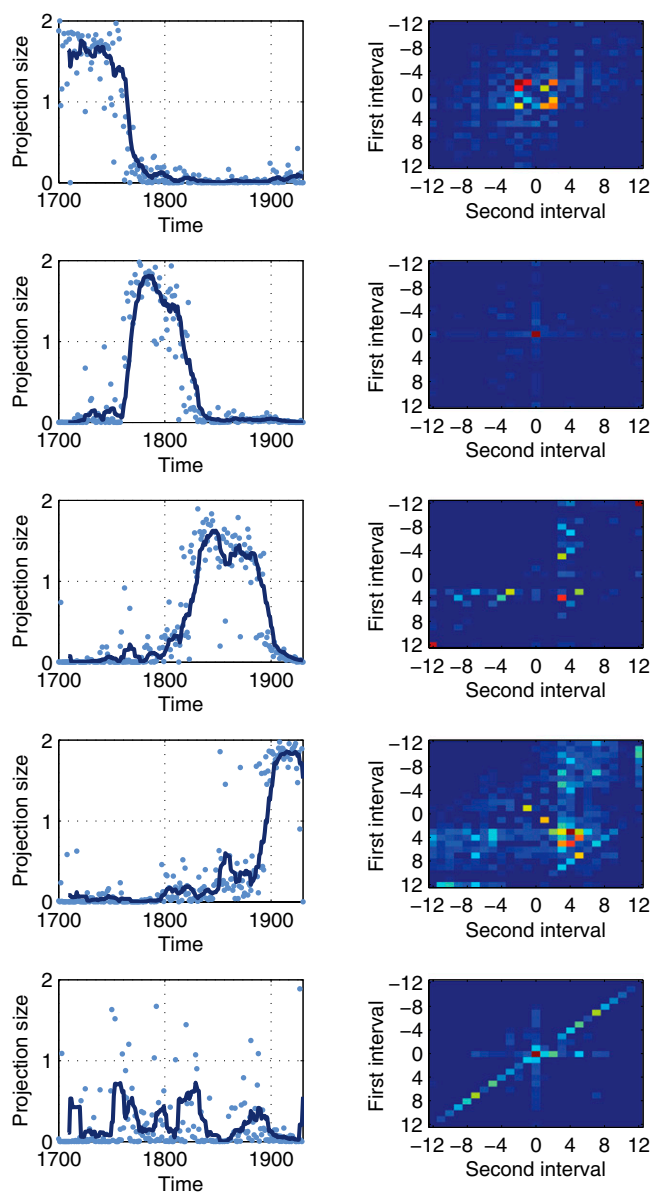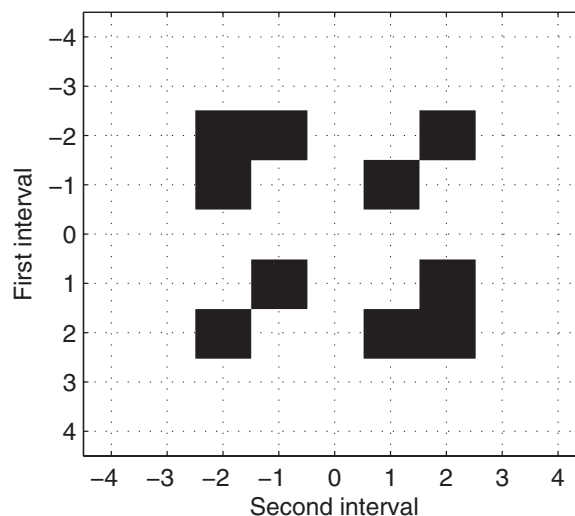
Fig. 3. Visualization of a matrix representing adjacent intervals in a scale. A cell is painted black if it is a bigram that can be built from three adjacent, also called diatonic, notes in a major/minor scale.

the probability distribution of melodic interval bigrams suggests that listener expectations may have changed as well. In consequence, we propose that a theory whose aim is to model listeners' expectations should be able to distinguish style transitions that changed event probabilities. Moreover, as conceptualized by Meyer (18), it is evident that composers seek to elicit emotions from the listeners by satisfying or denying their expectations. This may be reflected in how the IR principles are balanced, underlying stylistic changes as composers play with listeners' expectations. Thus, it is reasonable to expect that specific styles consist of certain fixed combinations of these principles.

In this context, we present a comparison with the IR theory, which has two properties that makes it relevant. First, it predicts on short-term music expectation, and, second, it is described as a set of principles that are formally coded as matrices. In the following paragraphs, we introduce the theory and then compare it with our results.

The IR theory (3) describes musical melodies as a succession of closures, implications, and realizations. Closure refers to a point in the melody at which expectancy for melodic continuation is weak.[‡] When a melodic interval, or interval for short, does not achieve perceptual closure, this interval, now termed implicative, generates an expectation as to what will follow, the realized interval. According to the theory, the listener's expectancies are ruled by a set of five principles, determined in large part by tone differences in the interval, such as size (small or large) or direction (increasing or decreasing). The principles are as follows:

- Registral direction: Small implicative intervals imply realized intervals in the same direction, but large implicative intervals imply realized intervals in an opposite direction.
- Intervallic difference: Small implicative intervals imply realized intervals that are comparable in size,[§] but large implicative intervals imply realized intervals that are smaller in size.[¶]
- Registral return: The difference between the first tone of the implicative interval and the second tone of the realized interval is no greater than two semitones.

Fig. 2. Recovered trends and factors of decomposition matrix *T*. (*Left*) Computed trends as a function of time. (*Right*) Matricial representation of the joint version of the factor whose trend is plotted to the left.

[‡]A complete description is provided by Carol and Krumhansl (21).

[§]Within a minor third if the registral direction of implicative and realized intervals is the same, and within a major second if the registral direction is different.

[¶]Smaller by more than a minor third if the registral direction is the same, and smaller by more than a major second if the registral direction is different.
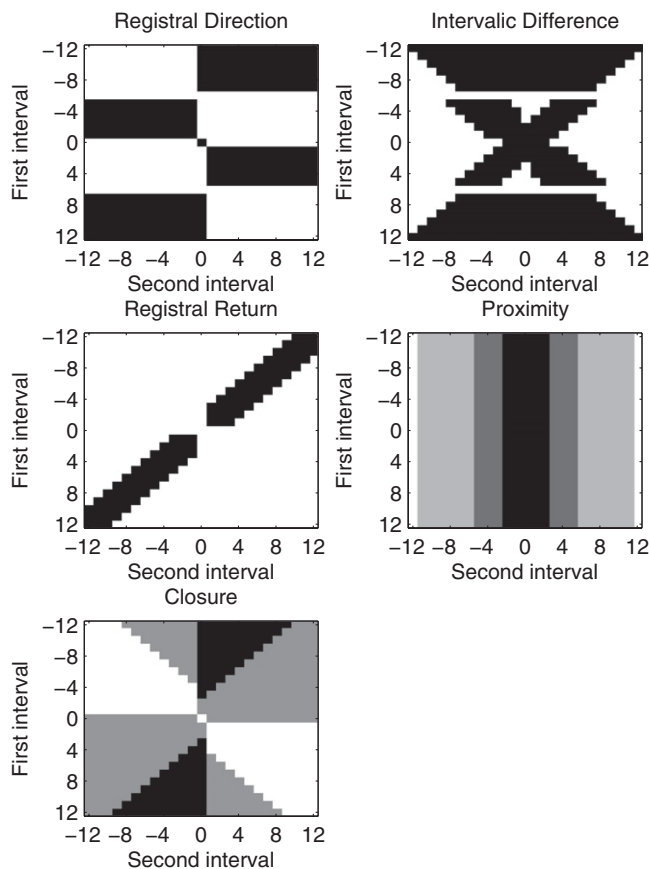
Fig. 4. Graphical display of the original IR principles. Each principle is coded as a matrix whose rows correspond to the size of the first melodic interval. Columns represent the size and relative direction of the second melodic interval. A cell in the matrix is interpreted as a score assigned to the bigram whose sum across all matrices corresponds to its expectability.

- Proximity: Independent of the size and direction of the implicative interval, implied realized intervals are no larger than five semitones.
- Closure: Closure is strongest when (*i*) the implicative interval is large and the realized interval is smaller[||] and (*ii*) registral direction of the implicative interval and realized intervals are different.[||]

A graphical representation of these principles is presented in Fig. 4. Each principle is coded as a matrix whose rows correspond to the size and direction of the implicative interval, and columns correspond to the realized interval. Finally, values in the matrix specify the assigned score to a specific pair of implicative and realized intervals: The higher the score is, the more expected is the realized interval. For example, the dark area depicted in the lower left corner of the matrix corresponding to the registral direction principle (Fig. 4) represents that intervals whose size is larger than six semitones are expected to be followed by any other interval of opposite direction.

Following Narmour's theory, expectations occur in relation to the degree of implication that the last heard interval generates. Hence, it would be difficult to compare the principles with the probability distributions because the latter were computed using all melodic intervals from all voices. This question has been addressed by Thompson and Stainton (23), who implemented an

algorithm to split implicative intervals from closural intervals. The expressive power of the IR principles was then compared, finding a negligible difference between the predictive power of IR theory in both implicative and closure contexts. In addition, the fact that the IR principles were able to predict continuations of actual music let the authors conclude that perception and production are guided by similar principles of expectancy.

Supporting this finding, there is also empirical evidence that Narmour's theory not only accounts for the variance of perceptual tasks but for production tasks (4, 5, 20).

Formally, we hypothesize that the IR principles can be regarded as a basis of melodic interval probability distribution space. Following this hypothesis, we projected data onto IR principles and analyzed their predictive power (*SI Materials and Methods*). Due to the fact that algebraic projection is equivalent to linear regression, each year's projection can be regarded as a regression using IR principles as predictors. We thus can measure each year's explained variance using R^2, resulting in a value of 0.23 ± 0.05. This
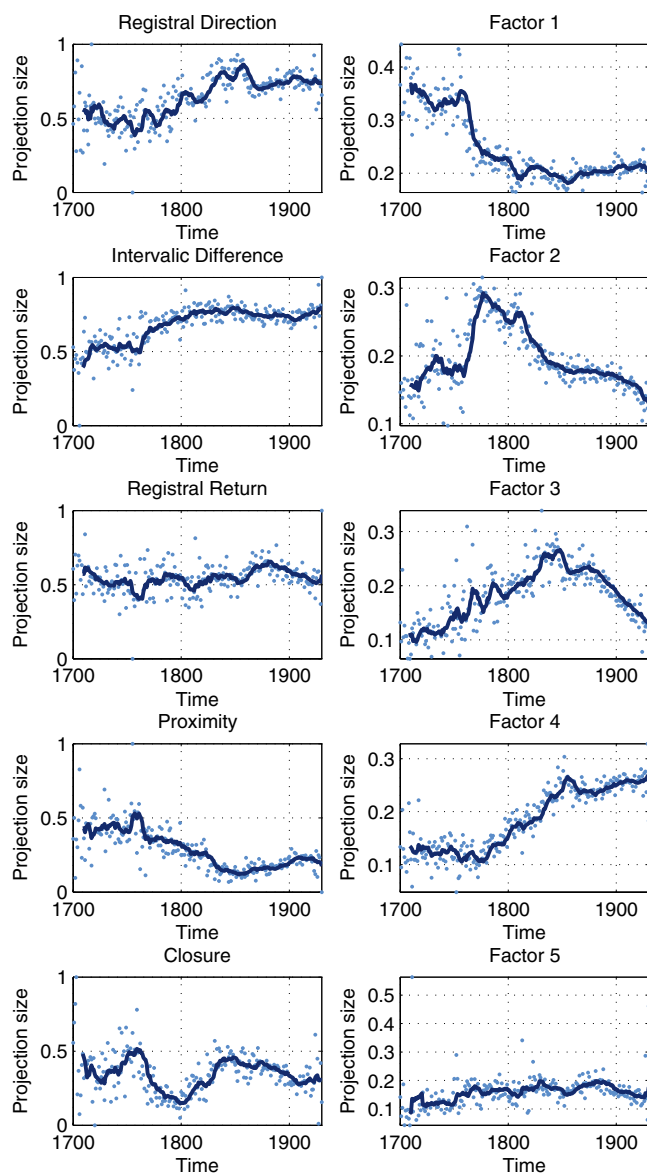

Fig. 5. Data projection over both IR principles and NMF factors. (*Right*) Projection of all IR principles. (*Left*) Projection of the NMF factors. Blue dots are the actual projection size, and the line represents a simple moving average.

[||]The description is taken from Carol and Krumhansl (21).

is a low value if we take into consideration the $0.98R^2$ of the factors of Fig. 2, suggesting that most of the variance of the empirical distribution is not captured by the IR principles yet trends might still be meaningful.

Fig. 5 (*Left*) depicts data projected over each IR principle. We analyzed the clustering power of IR principles by comparing it with both the results presented in Fig. 1 and random surrogates. Clustering results of IR projected data are qualitatively comparable to random surrogates; both yielded solutions with similar cluster sizes and the temporal continuity (details of surrogate construction and used metrics are provided in *SI Materials and Methods*).

Despite the fact that IR theory seems to be an unsuitable basis, some trends are still present. Fig. 5 (*Right*) includes the projection over the NMF factors. The projection of registral return is consistent with the fifth factor, and the registral direction principle with the fourth factor's projection. It also seems that Classical period has a lower closure than the rest of the periods.

Because IR theory has a relatively low predictive power, we propose that it can be boosted with data-driven factors, such as the one depicted in Fig. 3. This factor has already been used in the work of Thompson et al. (20), where the authors presented unfinished melodies to subjects, who rated candidate continuations, and then predicted their answers using IR theory and their own factors. The unison factor (Fig. 2, factor 1) was also used when predicting subjects' responses (21). However, it received a negative weight in linear regression, implying that continuations were worse when they ended in unison. This suggests that there might be a difference between production and perception tasks, because unison is very likely in scores but correlates negatively with human ratings of melodic continuation.

To summarize, we present a machine learning methodology that accurately identifies stylistic changes over the past two centuries. Furthermore, by means of matrix decomposition techniques, we found a set of four interpretable factors that explains this result. By analyzing these factors, we found structure that was both musically and cognitively relevant. Musicologically, we interpreted factors as key characteristics of Baroque, Classical, Romantic, and Post-Romantic music. Cognitively, we take advantage of the intrinsic relation between expectation and probability distributions, which allows us to interpret our findings within a psychological framework: the IR theory. In this context, we recovered two factors previously identified on purely intuitive grounds, diatonic (20) and unison (21), and two unique factors that can be considered as theoretical predictions requiring further psychophysical validation.

Materials and Methods

We present here a summarized description of the materials and methods used; detailed information is provided in *SI Materials and Methods*. As stated in *Results*, we selected the Peachnote corpus as the main material for this study (freely available at http://bochini.exp.dc.uba.ar/pub/prodriguez/raw_data.tgz). To compute reliable statistics, the corpus was cropped into a subset of contiguous years between 1700 and 1930. With the resulting subcorpus, we estimated the joint and conditional probability distributions of melodic intervals for each year y, formally $P_y(i_2|i_1)$ and $P_y(i_1, i_2)$. Taking advantage of the fact that melodic intervals are discrete, we flattened each distribution into a vector, which was then fed to a k-means algorithm (MATLAB, MathWorks). For orthogonal NMF, we used the MATLAB toolbox presented by Li and Ngom (24). After running the decomposition algorithm, the set of vectorial factors W was unflattened to recover a set of matrices depicted in Fig. 2. As a control (details are provided in *SI Materials and Methods*), we performed a similar analysis with pieces coded in standard Musical Instrument Digital Interface file format obtained from the Alicante 9 Genres Database (9GDB) corpus (25), which is also available for research purposes under request to the Pattern Recognition and Artificial Intelligence Group of the University of Alicante. This corpus contains a small number of works for the Baroque, Classical, and Romantic periods. Even though the works are in MIDI format, they correspond to a played version of them, which is a twofold control: On the one hand, it allows us to control the notational bias that may have been induced by the fact that the statistics were taken from scores; on the other hand, it allows us to test the generalizability of the factors found in the Peachnote corpus.

1. Sadie S, Tyrrell J, Levy M (2001) *The New Grove Dictionary of Music and Musicians* (Macmillan, New York), Vol 1084.
2. Serr J, Corral, Bogu M, Haro M, Arcos JL (2012) Measuring the evolution of contemporary western popular music. *Sci Rep* 2.
3. Narmour E (1990) The analysis and cognition of basic melodic structures: The implication-realization model (Univ. of Chicago Press, Chicago).
4. Carlsen J (1981) Some factors which influence melodic expectancy. *Psychomusicology: Music, Mind and Brain* 1(1):12–29.
5. Schellenberg GE (1996) Expectancy in melody: Tests of the implication-realization model. *Cognition* 58(1):75–125.
6. Cuddy L, Lunney C (1995) Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics* 57(4):451–462.
7. Viro V (2011) Peachnote: Music score search and analysis platform. *Proceedings of the International Society for Music Information Retrieval Conference* (Miami), pp 359–362.
8. Mitchell T (1997) *Machine Learning* (McGraw–Hill, New York).
9. Wertheimer, M. (1938) *Laws of Organization in Perceptual Forms* (Harcourt, Brace & Jovanovitch, London).
10. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755):788–791.
11. Ding C, He X, Simon HD (2005) On the equivalence of nonnegative matrix factorization and spectral clustering. *Proc SIAM Data Mining Conf* 4.
12. Li T, Ding C (2006) The relationships among various nonnegative matrix factorization methods for clustering. *Sixth International Conference on IEEE* (ICDM, 2006).
13. Choi S (2008) Algorithms for orthogonal nonnegative matrix factorization. *Neural Networks, 2008 (IEEE World Congress on Computational Intelligence)*, pp 1828–1832.
14. Bukofzer MF (1947) *Music in the Baroque Era: From Monteverdi to Bach* (WW Norton, New York), Vol 1084.
15. Simonton DK (1984) Melodic structure and note transition probabilities: A content analysis of 15,618 classical themes. *Psychol Music* 12(1):3–16.
16. Schoenberg A (1975) Style and Idea: Selected writings of Arnold Schoenberg. (Univ of California Press, Berkley), pp 216–217.
17. Huron DB (2006) *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge, MA).
18. Meyer LB (1956) *Emotion and Meaning in Music* (Univ of Chicago Press, Chicago).
19. Snyder B (2000) *Music and Memory: An Introduction* (MIT Press, Cambridge, MA).
20. Thompson WF, Cuddy LL, Plaus C (1997) Expectancies generated by melodic intervals: Evaluation of principles of melodic implication in a melody-completion task. *Percept Psychophys* 59(7):1069–1076.
21. Krumhansl CL (1995) Music psychology and music theory: Problems and prospects. *Music Theory Spectrum* 17(1):53–80.
22. Wertheimer M (1938) *Laws of Organization in Perceptual Forms* (Harcourt, Brace & Jovanovitch, London).
23. Thompson WF, Stainton M (1998) Expectancy in Bohemian folk song melodies: Evaluation of implicative principles for implicative and closural intervals. *Music Percept* 15:231–252.
24. Li Y, Ngom A (2011) The non-negative matrix factorization MATLAB toolbox for biological data mining. Technical Report 11-060 (School of Computer Science, University of Windsor, Windsor, Canada).
25. Prez-Sancho C, Rizo D, Iesta JM (2009) Genre classification using chords and stochastic language models. *Conn Sci* 21(23):145–159.