Associative Similarity **in Music**

From *A Fresco*: *Mélanges offertes au professeur Etienne Darbellay*, éd. Benno Boccadoro et Georges Starobinski (Bern : Peter Lang, 2013, 247-268. )

## Abstract

The accumulating evidence of perceptual studies in psychology, search experiments for audio data, and generative procedures for composition based on attention to close detail in the structure of existing musical repertories appears to converge towards a common point: much of what constitutes “similarity” to the listener lies outside the script of the music as notated. Here we examine five tangents that move towards a common hub of what we call associative similarity.

In all music-related disciplines, the most common focus of similarity studies has been on specific features of music—interval sequences, rhythmic patterns, harmonic plans, and combinations of any two of them. After briefly reviewing some of the recent approaches of this kind, we focus our attention on perceptual cues that are related to phenomena relevant to the performed (that is sounded) music but not explicitly present in the underlying composition as represented in writing. Under this second rubric, which we call associative similarity, we look at stylistic, timbral, and expressive similarity.

## Introduction (Feature Similarity)

Discussions of similarity are frequently based on a close reading of musical features—that is, elements of musical content, such as profiles of pitch movement, duration, and dynamics. In some cases, the study in based on audio content, in which case tempo, timbre, and texture may receive significant notice. In other cases, the study is based on musical notation or music-theoretic concepts (scale, mode, tonality). In experimental psychology, most studies related to similarity concentrate on artificial (lab-composed) examples which are short, lack variety in tempo or timbre, and are notably short. In examining the outcomes of studies conducted on the basis of such different presentations of “music,” it is inevitable that some findings can be attributed to these phenomenally different kinds of content. Unfortunately, the same divergences of meaning occur in the scholarly use of the terms “similarity” and “musical feature.”

The search for systematic approaches to feature similarity has a formal history that now stretches to more than 40 years. In the nineteenth century vast hordes of “tonic sol-fa” encodings of tens of thousands of hymn tunes enabled manual searching and sorting routines that were easily implemented in computational environments. Although automatic searches are much faster than manual ones, the algorithms involved are little changed. Tonic sol-fa codes (taking into account pitch, duration, meter, text syllables, and so forth) proved sufficiently durable to support the widespread circulation of pedagogical material, parlor songs, and show tunes in the nineteenth century, particularly in the British Isles.

Meanwhile on the European continent, another code was developed to describe folk songs in their monophonic instantiations. This code (the descendant of which is most widely known today as Essen Associative Code, or EsAC) likewise indicated pitch, duration, meter, key, scale, mode, the social context to which each song appertained, and (something not strictly present in a notated score) the phrase structure (as inferred from lyrics). Hundreds of thousands of folksongs which were transcribed or manually encoded in the nineteenth and early twentieth centuries used this code. They lie largely neglected in archives throughout Germany, Poland, Hungary, the Balkan Republics, Benelux, and elsewhere. Unlike tonic sol-fa repertories, which are unexplored as a basis for modern projects, EsAC code has provided the basis for the earliest software for automatic, feature-based assessments of similarity. Programs by Wolfram Steinbeck (1982), the late Helmut Schaffrath (1982-1994), and various pupils of his have kept EsAC in constant use for the past 25 years.

Knowing what features are most important in a given repertory is a great aid to the analyst. This is particularly the case for similarity searching. Every repertory is most easily searched with feature-weightings suited to its most prominent characteristics. One of the most innovative tools developed for the EsAC databases took advantage of the tendency of traditional music to conceal variants from searches based on initial melodic material but to reveal them to phrase-ending profiles. That is, adaptations and paraphrases are most likely to occur in the first bars but to preserve the original rhyme scheme in lyrics and the original musical structure of the earliest exemplar. Their search routine grouped together songs which had matching final phrase tones and corresponding numbers of phrases. It succeeded where more even-handed “generalized” approaches would have failed. The Essen researchers could utilize this approach because their data structure embedded phrase boundaries (derived from lyrics). [[1]](#footnote-1)

It is in its contrast with more typical feature-searches that cadential profiles stand out.

Extracted information that links content with semantic understanding (in this case cadential tones) is a harbinger of some of the work cited below, for the linking of detail with its understanding appears to play a central, but poorly understood, role in identifying what is, or may be perceived to be, similar.

1. Feature similarity (Lartillot/Toiviainen: standard stuff)

The approaches we discuss below concern stylistic, timbral, “mood,” expressive and expressive similarity. They draw selectively on recent work for the sake of illustration. No treatment is other than cursory, though important questions inhere in each.

Stylistic similarity (Cope: artificial music—all; broad-based)

Timbral similarity (Ebcioglu, QQ: artificial music—harmonization)

“Mood” similarity (Pachet: recordings; very general audio approach)

“Intensity” similarity (Eitan/Granot: artificial samples)

Stylistic similarity, while accessible to perception, is also accessible to logical thought and mathematical manipulation. Timbral similarity can be suggested by verbal cues on a written score, but makes its greatest impression through perceived sound. “Mood” similarity is highly dependent on styles of performance and their expressive qualities. “Intensity” similarity is concerned with the perception of change (of tempo, of dynamics, of note durations) in performed music.

The materials relevant to our enquiries consist of artificially composed and/or enhanced compositions in the first two cases, performed and/or recorded music in the first three cases, and artificially constructed examples in the last case.

## Aspects of Associative Similarity

### 1. Stylistic similarity (Deliege, Cope: artificial music—all; broad-based)

One approach to the assessment of similarity that is overlooked in music-theoretic and cognitive studies alike is that employed in the composition software developed by David Cope (University of California at Santa Cruz) between 1980 and 2005. Cope’s Experiments in Musical Intelligence (Emmy), an elaborate set of procedures designed to synthesize new pieces from old, generated thousands of new compositions based on a multifaceted analysis of exemplars by particular composers from specific genres within the repertory of each. It had three components—a lexicon for storing analytical information, a grammar with five “parts of speech” used to generate all compositions, and an interface designed to accommodate user preferences for genre, composer, key, meter, mode, and other features. The software reproduced such a wide range of different kinds of music that a representative list reads like an undergraduate program in the study of repertories. From the Western classical tradition it included Renaissance polyphony, Bach inventions and Brandenburg concertos, Mozart piano sonatas, flute concertos in the style of C. P. E. Bach, piano pieces in the style of Rachmaninoff and Prokofiev, an opera “in the style of Mahler” [Mahler wrote no operas], and much else. Given all those models, it is surprising that the same process produced simulations of Indonesian gamelan music (as well as synthesis of Indonesian and European classical music) and “Broadway” songs (complete with synthesized lyrics) in the style of Jerome Kern [xx]. Most of the pieces it generated could be recognized as impressive (if not always persuasive) imitations of their models. (Every piece was synthesized from isolated features of at least two genuine [humanly-composed] pieces.)

In this context “similarity” refers to the ability to mimic a pre-existing model without replicating surface details in a literal or overly obvious way. (Cope argues that Emmy does truly “compose” music and that what is mimicked is the *process* of composition rather than prior compositions *per se*. For the purposes of the present enquiry, I shall leave aside this endlessly stimulating debate.)

In 1997, when the Center for Computer Assisted Research in the Humanities hosted a series of panels on creativity and the arts moderated by Douglas Hofstadter, a special concert in which pieces composed by Cope’s Experiments in Musical Intelligence and by the composers whose works were automatically “consulted” in the software lexicon were intermixed. The works were performed live by a pianist and, where appropriate, a violinist. The program was not made available to the audience until the concert was finished. Attendees had been asked to judge whether each work was genuine or artificial. Although the idioms were familiar ones to an audience accustomed to classical music, none of the genuine works was particularly familiar. Inevitably there was more consensus on some works than on others, but in most cases opinion as to authenticity was divided in the 50% +/- 10% range. The audience consisted of roughly 200 people, the vast majority of whom participated in the evaluations. It consisted largely of adults, of which less than 50% were university students but nearly all of whom were highly musically literate.

It is useful in the present context to examine exactly what kind of information was available to those making these judgments. The distribution of genuine and synthesized works was randomly distributed in the sequence of works. The listeners all heard the same music performed in the same way at the same time. In the case of the synthesized works, the “genetic” content at the event level was limited to pitch and duration (via the MIDI data used to encode works for analysis and eventual resynthesis). However, this content was also modeled indirectly on higher-order (though poorly defined) cognitive phenomena such as “genre” and “composer.” “Genre” and “composer’s style” are ultimately filtered through semantic schemas that determine the actual sequence of elements used to fulfill the five-parts-of-speech grammar. (Numerous variants of some of the same “works” show that length is somewhat indeterminate.) The “elements” are not single events but rather short sequences of events (typically 2-5). When they are identified as “characteristic” sequences (often recurring but not unique to one piece), they are stored in the lexicon with information about “antecedents” and “consequents”—that is, clues to the context in which they occur.

To all of this, the act of performance confers a third level of features—tempo, volume, such elements of expression as each instrument enables and each player provides. In this particular context, these operate at the highest order of abstraction, since they are likely to change only by phrase or section**. I shall return to** this aspect of the judgments further on. xx

### 2. Timbral similarity (Ebcioglu; QQ; artificial music—harmonization)

The frustration of timbral expectation has been noted again and again to be confound listening activities, at least in some subjects. Systematic studies from the realm of psychology are elusive. (br appears to be the only thing that is somewhat relevant. On it, more later.)

Since 2002 we have had on line a game-like quiz which was intended to get at some questions of stylistic similar. Designed by Yi-wen Liu and implemented by Craig Stuart Sapp, the Haydn-Mozart Quartet Quiz presents users with a random movement from a string quartet by one composer or the other. Users rate their familiarity with this repertory on a scale of 1-5. Liu’s intent was to cull from the positive identifications clues to the traits that listeners use to differentiate two closely related repertories. His information-theoretic analysis eventually enabled a machine-driven stylistic discriminator to come within two percentage points of human subjects. Some clues were also gathered by interviewing those near to hand who had played with the Quartet Quiz.

The most unexpected finding concerning how users relate to such a task was that for a small percentage of users, the fact that MIDI strings hardly sound like strings at all (though this will vary from one user’s machine to the next) deafened the listener to other traits of the music. (Features sited as useful clues pointed to characteristic rhythmic patterns, textural contours, tessitura, harmonic change, and melodic figuration as generally useful.) How such traits might be prioritized seemed to vary from piece to piece. Clearly for the feature-deaf listeners, timbre was a primary consideration in stylistic estimation. We think that feature-deafness in the presence of an irritating “unnatural” timbre may affect about 5% of subjects. Since the finding was not anticipated at the outset, we compiled no statistics. Hadyn is correctly identified 57% of the time (16,002 trials), Mozart 55% (16,004 trials) of the time. For individual works, there are substantial differences in recognition. The current ranges for Haydn are 38% (Op. 1, No. 1, movement 5) - 79% (Op. 9, No. 3, movement 2). The Mozart they are 26% (K. 169, movement 3) – 74% (K. 590, movement 1). In other words, for those who can ignore “unnatural” timbres, Haydn more consistently sounds more “like Haydn” than Mozart does “like Mozart.” The database contains 196 quartet movements by Haydn and 80 by Mozart.

Although frustration of timbral expectation seems to have played only a small role in confusing Quartet-Quiz users, we had noticed previously that listeners could be so aware of timbre that they might be influenced to misjudge (or at least fail to observe) features of musical content generally regarded in the evolution of music theory to play a central role in defining style and genre. We have long used Kemal Ebcioglu’s artificial harmonizations of Bach chorales in class to demonstrate the pros and cons of rule-based approaches to such activities. Ebcioglu found (1986) that ubiquitous though rule systems for four-part harmonization might be in the literature of music theory, the total harvest of written rules was vastly insufficient to generate credible harmonizations. Eventually he produced 300 rules. In implementing the rules, his program needed to “look” forward and backward, just like a student in a harmony class, to make sure that each note added to the “score” did not violate a contextual rule, such as the prohibition of parallel octaves and fifths. The system seemed to lack gestural information (what combinations of notes can human hands actually play at one time) and was prone to space notes rather widely apart across three octaves. There was also less rhythmic differentiation between voices than would be found in many authentic harmonizations.

However, when examples from “real” and “pseudo” Bach harmonizations are interspersed, and when each is coupled with a constantly changing range of (electronically controlled) temperaments, many listeners seem to consistently identify a chorale played in mean-tone temperament as being “Bach,” whether or not the harmonization is artificial. Conversely, harmonizations played in equal temperament are often judged to be artificially created. It seems to take a well-trained ear to ignore temperament in favor of voice-leading and composite rhythmic activity.

In a series of public symposia featuring Douglas Hofstadter and David Cope (1997), we provided a concert in which actual pieces from classical repertory were intermixed with pieces composed by Cope’s Experiments in Musical Intelligence Software. The EMIS generated new works in which an algorithm based on a modified augmented transition network dipped into a database of features segregated by composer, genre, movement, and other rubrics to “fill in” a template which had to use each of its gleanings in a manner that grammatically satisfied the situation of the fragment in a humanly-composed work. There were ten works on the program. All of them were performed live by two performers—a pianist and a violinist. Although the “real” works were all by well-known composers, they were chosen on the basis of the likelihood that they are rarely played and would not be familiar to most of the audience, which numbered about 200. The composers’ names were withheld from the program. Attendees were asked to judge whether each work was real or artificial. The range of responses somewhat resembled those of the Quartet Quiz, although now the timbres were entirely “natural.”

This anecdotal result conforms to the general experience listeners living near the University of California at Santa Cruz, where Cope has long been a professor of music. EMIS compositions in a much broader range of styles and genres (including Renaissance polyphony, nineteenth-century opera, and twentieth-century piano music) have been performed live at numerous concerts by appropriate performing groups. The EMIS repertories, which eventually came to include “songs from Broadway musicals” in which both the lyrics and the music were concocted by EMIS, became familiar to those frequently the concerts. People eventually culled their favorites. Cope, on the other hand, destroyed thousands of “outputs” that did not seem to produce compelling works. The point is, once again, that timbral similarity seems to be a dominant features in identification and recognition of style and genre.

### 3. “Mood” similarity” (Pachet: recordings; very general audio approach)

Much recent work of Jean-Julien Aucouturier and François Pachet (ap) and various colleagues at Sony Labs in Paris has been devoted to something that seems well beyond the bounds of systematic studies of listening. It is focused on “recommendation services”, that is methods for scoring works according on several yardsticks aimed at facilitating the identification of new works that will please the listener because they are “similar” to a previously identified preferred work. Recommendation services are now legion, and it is inevitable that any two will share something in common. However, Pachet’s work was essential to building the ground floor of this enterprise. The procedures are, of course, intended to operate automatically.

Aucouturier and Pachet reject “collaborative labeling” because it is entirely divorced from musical content. In approaches based on the extraction of audio signals (for example, of tempo, rhythm, or melody) they note the difficulty that untrained listeners have in identifying specific musical features. The avenue they pursue is that of timbral quality. They note, in passing, that pinning down melodic similarity is an elusive goal because the change of a single “can dramatically impact the way it is perceived.” (91) In contrast, small changes in timbre do not seem prone to distance one example from another.

As part of the CUIDADO European IST Project, they created a database of 17,075 popular pieces. For processing timbral information they used a Mel Frequency Cepstrum Coefficient (following Tzanetakis 2001). A typical three-minute piece would be represented by 3600 feature vectors. Then every set would be compared to every other set. Subsets of candidates were culled by searching for (a) multiple recordings of the same title, (b) songs sung by the same artist, (c) performing medium, and (c) songs in the same genre.

What is spectacular about their results is that sector which juxtaposes quite unlikely pairs that do, indeed, seem to embed common timbral qualities that would never on the basis of scripted content be noticed. Take, for example, Beethoven’s “Romanze für Violine und Orchester Nr. 2,” Op. 50. It showed a high correlation with both “Singing in the Rain” (sung by Gene Kelly) and the Beatles’ “Eleanor Rigby.” A recording by the Orchestre Symphonique de Montreux of Gershwin’s *Porgy and Bess* had much in common timbrally with the first movement of Prokofiev’s Fifth Symphony, Op. 100 (*Celibidache*). Robert Schumann’s “Kreisleriana,” Op. 16, No. 5 (*Sehr langsam*), played by Horowitz, was highly correlated with Bill Evans’ “I love you Porgy.”

Among the more perfunctory results were that different performances of the same title almost always had a high correlation, while different pieces by the same composer or in the same genre did not necessarily have such a correlation. What is also important about this work in the search community is that pairs with high timbral correlations were rarely suggested by clustering of metadata. In short, labels do not tell us much about the aural impression recorded music will make.

### **4. “Intensity” similarity** (Eitan/Granot: artificial samples)

Eitan and Granot have run a series of experiments to study ways in which the brain appears to confuse an experimental feature with a control feature when the first is intensified. The manipulated features include such variables as pitch which rises or falls, volume which increases or decreases (crescendo, diminuendo), and tempo which accelerates or decelerates. Users appears to map the overall impression of change to a sense of spatial motion. For example, a series of ascending pitches may be co-perceived as a spatial rise or as accelerating, even when the tempo is unvarying. Differences between trained and untrained subjects were small. The researchers suggest that there is some interplay between “tension” and content.

Their findings suggest, at the very least, that the analysis of isolatable musical features, however continuous in complete works, cannot reveal the dynamic processes at play in listening experiences. They question whether the overall results are more suggestive of a congruence or a compensation hypothesis. That is, Do intensity and the musical feature to which it is cross-attributed enhance one another, or, If there is more of one is there less of the other? Despite a survey of much recent literature that is of tangential interest to these questions, they propose a continuing series of experiments to explore both possibilities systematically, and always with the emphasis on perception.

The authors acknowledge among the literature that broaches such coupling anomalies an article on ramp archetypes (Huron 1992). Its purview is quite limited, however: it investigates the use of crescendos and diminuendos in selected works by 14 composers and finds that diminution takes place quickly relative to time spent “ramping up” to a fortissimo. More to the general point is Huron’s *Expectation in Music* (2006), which delves deeply into an array of associative factors in music response. Huron posits that many perceptions and reactions to musical phenomenon reflect animalistic interpretations of specific sound phenomena which induce visceral responses necessary to promote survival of the species. These are often relationships that must remain conjectural, because there is no accurate way to measure the totality of possible reactions to screeches, screams, hoots, and the like. However, music that mimics such auditory phenomena, and certain physiological responses to it, can be measured under controlled settings.

For the purposes of the present enquiry, the general thrust of Huron’s work may be relevant. Why should human listeners respond to one attribute of music as though it were another? What purpose does it serve?

Eitan and Granot present a still more unlikely finding in their 2008 installment of their ongoing experiments. It involves the use of true and (at least partly) false compositions, reminiscent of Cope’s Experiments, in order to explore the subject of whether music-theoretic ideas of thematic coherence (along the lines of Rudolf Reti’s postulates) can be verified in controlled experiments. The short answer seems to be no, or at least not convincingly so.

Challenges the notion of “organic unity” of large architectures of “musical form” by determining that listeners are not bothered by episodic “realizations” of sonata form in which extraneous sections are substituted in recapitulations. Neither trained not untrained subjects, listening to a hybrid work once, or several times, varied significantly in their indifference to “coherence.” Mozart’s sonstas K. 280 and 332 were used as seed material. In fact, in the case of the K. 280 and its hybrid the majority of subjects preferred the hybrid.

## Discussion

Despite the fact that all these approaches issue from distinctly different disciplinary quarters, they all assert some claim to better measuring and/or understanding similarity judgments.

(br) Brancucci, Alfredo, and Pietro San Martini, “Hemispheric Asymmetries in the Perception of Rapid (Timbral) and Slow (Non-timbral) Amplitude Fluctuations of Complex Tones,” *Neuropsychology* 17/3 (2003), 451-457.

Suggest the importance of the right hemisphere in perceiving temporal variations of intensity whether or not the rate of variation is rapid (perceived as timbral variation) or slow (perceived as dynamic change). (The subjects were non-musicians.)

Sources cited

Aucouturier, Jean-Julien, and François Pachet (2002). “Finding Songs that Sound the Same,” *Proc. 1st IEEE Benelux Workshop on Model-based Processing and Coding of Audio (MPCA-2002)*, Leuven, Belgium. Nov. 15, 2002.

Brancucci, Alfredo, and Pietro San Martini (2003). “Hemispheric Asymmetries in the Perception of Rapid (Timbral) and Slow (Non-timbral) Amplitude Fluctuations of Complex Tones,” *Neuropsychology* 17/3, 451-457.

Cope, David (2001). *The Virtual Score: Computer Synthesis of Musical Style.* Cambridge: The MIT Press.

Eitan, Zohar, and Roni Y. Granot (2007). “Intensity Changes and Perceived Similarity: Inter-Parametric Analogies,” *Musicae Scientiae*, Discussion Forum 4A, 99-133.

Eitan, Zohar, and Roni Y. Granot (2008). “Growing Oranges on Mozart’s Apple Tree: Inner Form and Aesthetic Judgment,” *Music Perception* 25, 397-417.

Huron, David (1992). “The ramp archetype and the maintenance of passive auditory attention,” Music Perception 10, 83-92.

Huron, David (2006). *Expectation in Music*. Cambridge, MA: The MIT Press.

Liu, Yi-wen, and Craig Stuart Sapp (2002). “The Haydn-Mozart Quartet Quiz”. <http://qq.themefinder.org>.

Schaffrath, Helmut (xx).

Steinbeck, Wolfram (1982). *Struktur und Ähnlkichkeit: Methoden automatisierter Melodienanalyse* (*Kieler Schriften zur Musikwissenschaft*, 25)*.* Kassel: Bärenreiter Verlag.

Tzanetakis, George (2001). “Automatic Musical Genre Classification of Audio Signals,” *Proc. ISMIR*.

Zharipov, xx (1968).

The only common denominator among studies of similarity is that they all involve (at least potentially) the act of listening. It too throws up roadblocks to a common understanding, for numerous people taking a cognitive approach allocate to listening, rather than to musical content as defined by a score or a recording, the full weight of similarity judgments. For some, musical content seems to be irrelevant. Two pieces, or melodies, are the same (or different) if the subject says they are. There is clearly a danger of total subjectivity in this attitude, and yet, as similarity studies continue their rapid evolution under diverse auspices, there is some support for this view.

1. The same approach has been used by John Walter Hill to trace text paraphrases in early Roman declamatory music (related to seventeenth-century opera) and to identify the retexting of opera arias in the Vivaldi repertory. His approach is based on scansion-sensitive text-search software. The Roman *seicento* works incorporate the same tendency as the traditional songs to vary beginnings but to preserve cadence patterns and phrase structure. [↑](#footnote-ref-1)