

Associative Aspects of Perceived Musical Similarity and their Intersections with *Seconda-prattica affetti*

Abstract

What constitutes musical similarity? The range of possible answers is a blinding array of choices, all of which make certain assumptions. They are variously founded on written notation, constructs of music theory, cognitive theory, and many other things. They favor the views of performers, scholars, or “listeners” according to the origin of the research. Here we single out one thin slice of this array—recent audio and perceptual studies—and seek parallels to precepts of the early Baroque, particularly Claudio Monteverdi’s doctrine of the affects.

1. Associative Similarity

Under the rubric of *associative similarity*, we comment on (1) feature-based descriptions of compositional style; (2) timbral aspects of similarity judgments, (3) similarities of musical “mood” as identified by listeners, and (4) influences of musical “intensity” on similarity judgments.

1A. Feature-based similarity

The Lexicons of David Cope’s Experiments in Musical Intelligence

The word association assumes the presence of a relationship. In music, one work can be associated with another through an endless variety of traits—its genre, its composer, its chronological period, its place of origin, the cultural heritage it represents and so forth. When one delves into the more finite world of musical style, no one has focused so much attention on traits that define the “style” of one genre, composer, et al. than David Cope, the originator of a project in artificial composition called Experiments in Musical Intelligence (Emmy for short; see 1991, 1996, 1998, and 2006), recently succeeded by Emily Howell, a virtual composer. The “experiments” began in c. 1980 and continued for 25 years, over the course of which many extensions and refinements were applied to the system. It consisted mainly of two elements—a lexicon and a grammar. The *lexicon*, which contained separate compartments for specific genres within the oeuvre of individual composers of many times and places, absorbed, through recursive processing of thousands of MIDI files, tens of thousands of style quotations—short (usually polyphonic) segments of actual pieces. In processing approach and departure details for each segment were also stored.

The grammar identified five functional components of a satisfactory movement or single-movement work. These parts were a statement, preparation, extension, antecedent, and consequent. A user of the system could select a composer and a genre, then provide up to 15 global variables (key, mode, meter, etc.) to use in the generation of a virtual piece in the defined style. After the piece was generated, the user was at

liberty to save it or delete it. Most of the thousands of works generated by Emmy do not survive, but those that do in many instances provide striking caricatures of the works whose DNA they contain.

Emmy's early lexicons were based mainly on keyboard music (Bach inventions, Mozart sonata movements, Schumann, Prokofiev, and so forth). To them were gradually added Palestrina masses, Indonesian (especially gamelan) music, Broadway songs, and eventually operas. In its final ten years of activity, Emmy began to provide lyrics for some of its repertoires including "Gershwin" songs and the opera *Mahler* (with textual material acquired from the letters of Alma Mahler). The generation of texts eventually led to a new sphere of activity in which new lexicons (for words, phrases, and their musical analogues) could be developed, through human-computer conversation, to support the creation of complete works without the use of a pre-established lexicon. Thus was the virtual composer Emily Howell (debuted in 2010) launched.

That Emmy was able to operate across such musically diverse repertoires with a single grammar remains one of its most notable revelations about musical style. Conventional approaches to musical style do not posit the separation of grammar and lexicon; they seek instead to regard compositional substance as a transit from event to event, theme to theme, section to section, movement to movement. Abstract concepts such as "theme" are absent from Emmy.

These thoughts are worth remembering when confronting conventional notions of musical similarity. To someone unfamiliar with Baroque keyboard music, one Bach invention sounds much like any other Bach invention. For such a person, they could be said to be similar. Yet for a Baroque harpsichordist, no one sounds like any other. This distinction points to the obvious point that all notions of similarity are relative to the musical exposure of the perceiving subject.

That much being said, Emmy's criteria for identifying material appropriate for storage in a lexicon also provide what may be an important method for recognizing highly concentrated elements of a composer's discourse. These are "signatures" and "earmarks"—isolatable passages that are unique to a composer and sometimes a genre and to a single work or works written in rapid succession to one another. At their most generic level spread across the composer's oeuvre. However, they are discarded in the process of identification if they are so frequent as to be ubiquitous. Signatures typically consist of 2 to 5 consecutive events. Signatures stored in lexicons offer the strongest possibility of tracing content similarity but also genre differentiation. Mozart may use one figure over and over in his piano sonatas, but an entirely different one may be characteristic in his symphonies. Those used in Beethoven's symphonies will be nothing like those used by Mozart, which implies that genre alone cannot account for most traits of musical style. Genre will be defined more readily by the grammatical details that generate a characteristic structure. Despite the strong identification in Western art music of works by their key, mode, and genre (e.g. Concerto in D Major for Violin and Orchestra), Emmy testifies that user choice—of meter, mode, et al.—contributes only to superficial details of resulting compositions.

<earmark example>

Folksong families in the Essen Hochschule Projects

Knowing *a priori* what features most uniquely typify a repertory is a great aid to the folksong analyst. In the extensive projects based at the Essen Hochschule für Musik, mainly under the direction of Helmut Schaffrath, “similarity” occurred in some unlikely places. Large numbers of relatively simple songs collected by the Deutsche Volksliedarchiv (established in 1914),¹ having been subjected to many quantitative measures after mainframe encoding in the 1970s, were in the 1980s processed on personal computers provided with an ever-growing number of programs that sought to produce profiles that could enable the identification of clusters of works with common melodic and/or rhythmic substance. Much of digital humanities focus in the 1960s into the 1990s leaned towards first-line indices as valid tools for identification of textual content. Folksong research had long recognized that melodic (and textual) beginnings were more prone to modification in transmission than were endings. They also knew that modal content was rarely altered.

From these two analogue-era principles they easily developed new means of corraling variants on a single tune. One procedure generated a list of phrase finals by scale degree. The other gave (also by scale degree) a complete index of all the tones used in a piece. No folding of scale degree was permitted. If a piece had 17 different pitches, then it was listed as having a 17-note index of tones. If it used only six pitches, then it was classified as hexatonic. From these two procedures, many tune families could be identified in a relatively short time. The Essen researchers could employ the first procedure only because their data structure embedded phrase boundaries (generally derived from lyrics).² In the case of these highly diverse musical enquiries, a high degree of familiarity with the repertory being processed is fundamental to the development of appropriate techniques.

1B. Similarity of Timbre

Anyone who has used artificially-created sound material in controlled experiments will have noticed that some listeners foreground timbre to a degree that entirely eclipses their perceptions of other musical features—stylistic traits, melodic identity, harmonic structure, and so forth.

¹ Though unrelated to the topic of this essay, the rich offerings of the Deutsche Volksliedarchiv are well represented by their Liederlexikon (<http://www.liederlexikon.de>), which provides in-depth histories (with graphical display of text and music variants) of individual songs.

² 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.

For respondents to the Haydn-Mozart Quartet Quiz (2002) that has been served online by the Center for Computer Assisted Research in the Humanities in 2002, users are unable to select a desired timbre. They are presented with a MIDI-file and are asked to judge whether the movement being presented is by Haydn or Mozart. The experiment was designed (by Yi-Wen Liu) to compare machine learning of musical style with user judgments. As implemented by Craig Stuart Sapp, the Haydn-Mozart Quartet Quiz presents users with a random movement from a string quartet by one of the two composers. Users rate their familiarity with this repertory on a scale of 1-5. Liu's information-theoretic analysis eventually enabled a machine-driven stylistic discriminator to come within two percentage points of the accuracy of human subjects.

Unexpectedly, a small percentage of early users objected to the fact that MIDI files were used (the criticism has persisted). They would produce higher scores, they said, if they heard recordings of real performances. According to the originators, features usually sited as useful clues to composer identity were rhythmic patterns, textural contours, tessitura, presence or absence of abrupt harmonic change, and details of melodic figuration. How such traits might be prioritized seemed to vary from piece to piece. For timbre-sensitive users, however, the notoriously unconvincing simulation of string timbres by MIDI was such a big distraction that they were unable to attend to musical features *per se*. In psychological terms, an unexpected timbre constitutes a cognitive distraction. In Oakes (2006) slow tempos were correlated with higher levels of recall for associated advertising material. Patel et al (2008) consider these distraction effects the result of interference between syntactic processing and musical and/or linguistic inputs. Work in this area continues vigorously at the present time.

1C. Similarity of “Mood”

Were it not for our experiences with the Haydn-Mozart Quartet Quiz, we might once have been skeptical of the work of Jean-Julien Aucouturier and François Pachet at Sony Labs in Paris, because its notion of musical similarity is so distant from music-theoretic ones. Although it is an over-simplification of their work to say that they were looking for concordances of “mood”, it is not too far from the truth. Their motivation was to appraise musical works for a “recommendation service”—a commercial approach to suggesting what else users picking one recording might also like. We believe, though, that the Aucouturier-Pachet work exhibits results so striking that they raise profound questions for those pursuing definitions of musical similarity.

In the welter of music-and-social-networking initiatives that have sprung the first Aucouturier-Pachet publication (2002), their project is notable for having rejected “collaborative labeling” (crowd-sourcing) because it is entirely divorced from musical content as we general regard it. They note that pinning down melodic similarity is an elusive goal because the change of a single note (in pitch or duration) “can dramatically impact the way it is perceived,” even though in statistical data the difference may be barely noticeable. In comparing the possible value of extracting tempo, rhythm, or melody from audio material, they note the difficulty that untrained listeners have in identifying specific musical features. They

concentrate on timbral quality and place it under the rubric of “mood.” (They remark that small changes in timbre do not seem prone to distance one example from another in the responses of listeners.)

Under the auspices of the CUIDADO European IST Project they created a database of 17,075 recorded pieces. For processing timbral information they used a Mel Frequency Cepstrum Coefficient (MFCC, following Tzanetakis et al. 2001). A typical three-minute piece would be represented by 3,600 feature vectors. Every set of vectors would be compared to every other set. Clusters of “similar” candidates were culled by searching for (a) multiple recordings of the same title, (b) songs sung by the same artist, (c) songs involving the same performing medium, and (c) songs in the same genre.

What is striking about their results is the unlikely audio pairs it produces (Table 1).

	Sample	Work rated “similar”
<i>Same artist/piano</i>	<i>Franz Schubert: Op. 90, N 2 in Eb Major</i>	<i>Franz Schubert: Op 90, No. 4 in Ab Major</i>
<i>Same artist/harpsichord</i>	<i>JS Bach: Well-Tempered Clavier (WTC), Fugue in C Minor (BWV 847)</i>	<i>JS Bach: WTC, Prelude in C# Minor (BWV 849)</i>
<i>Same genre/piano</i>	<i>Mozart: Piano Sonata K. 533, movement 1</i>	<i>Scriabin: Piano Sonata No. 2; Weber: Piano Sonata Op 49, No. 3</i>
<i>Same genre/harpsichord</i>	<i>Bach: WTC, Prelude in C# Minor (BWV 849)</i>	<i>Couperin: Gavotte</i>
<i>Same orchestral textures</i>	<i>Wagner: Ride of the Valkyries</i>	<i>Wagner: Brunhilde</i>

Table 1. Various pairings of performing medium and genre with sample candidate matches from different classical repertoires.

Why would Beethoven’s “Romance for Violin and Orchestra”, Op. 50, No. 2, for example, show a high correlation with the Hollywood musical theme-song “Singing in the Rain”, or the Beatles’ “Eleanor Rigby”? In such cases there are no obvious matches between titles, artists, performing media, or genres. What might a recording of Gershwin’s *Porgy and Bess* have in common with the first movement of Prokofiev’s Fifth Symphony, Op. 100? Or Robert Schumann’s “Kreisleriana” for piano, Op. 16, No. 5 (*Sehr langsam*), played by Vladimir Horowitz, with Bill Evans’ rendition of “I love you, Porgy” (again from *Porgy and Bess*)? These culls suggest that particularities of performance superimpose their own imprint on the perception of content. Might common elements of *expression* form the essence of what likens one piece to another in the listener’s mind, even when differences of content and purpose are enormous? Does expression as provided by performers eclipse expression that inheres with the composition? These are pivotal points.

Somewhat in sympathy with Cope’s musical mash-ups (such as “Mozart in Bali”, in which Emmy-Mozart morphs into Emmy-gamelan music and the into a hybrid of the two), the Aucouturier-Pachet models identify specific complexes of acoustically similar features (comprising tempo, dynamic level, sonority as

defined mainly by timbre, et al.) in works from diverse realms of musical experience. In some cases they pair multiple variables as controls to explore a third feature. A few unexpected results (classified as “interesting” because of their anomalies) are shown in Table 2.

Performing medium	Genre contrast	Classical title	Other title
<i>Piano</i>	<i>Classical-contemporary</i>	<i>Rachmaninov: Moment Musical Op 16, No. 2</i>	<i>Ligeti: Concerto for Piano and Orchestra</i>
<i>Piano</i>	<i>Classical-jazz</i>	<i>Schumann: Kreisleriana Op 16, No. 5</i>	<i>Gershwin: “I loves you, Porgy” (Porgy and Bess)</i>
<i>Orchestral</i>	<i>Classical-jazz</i>	<i>Prokofiev: Symphonie Op 100, No. 5, movement 1</i>	<i>Gershwin: Porgy and Bess</i>
<i>Orchestral</i>	<i>Classical-pop</i>	<i>Beethoven: Romance for Violin and Orchestra Op 50, No. 2</i>	<i>Beatles: “Eleanor Rigby”</i>
	<i>Classical-musical</i>	<i>Beethoven: Romance for Violin and Orchestra Op 50, No. 2</i>	<i>Nacio Herb Brown: “Singin’ in the Rain”</i>

Table 2. Anomalous listener-defined similarities among works of different genres in the Aucouturier-Pachet research.

The Aucouturier-Pachet work does not ignore or discard more usual notions of similarity grounded in content: different performances of the same title almost always produce a high correlation. But correlations obtained in this context do not preclude the occurrence of associations between works from radically different contexts.

In contrast also to many early projects motivated by the quest for valid methods of audio recommendation” is the total irrelevance to the Aucouturier-Pachet model of parameters that can easily be clustered in metadata searches (for key, composer, date, etc.). This stands in partial contrast to Cope’s work, for different pieces by the same composer or in the same genre do not necessarily produce high correlations. *Timbre* can play a central role in assessing similarity and difference. That the Beatles’s songs “Helter Skelter” and “Lucy in the Sky with Diamonds” do not engender audio impressions of musical similarity is a finding that the authors attribute to differences of “orchestration”: the first involves “heavily overloaded guitars,” the second “tremolo organ.” One concludes that customary labels and content classification methods do not provide valid guidance on the aural impression of musical similarity that comparative listening seems to produce.

1D. Similarity of Musical “Intensity”

Zohar Eitan and Roni Y. Granot (2007) 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 rising and falling pitch, increasing and decreasing volume (crescendo, diminuendo), accelerating and decelerating tempo, and so forth. Users appear to map the overall impression of change to a sense of *spatial* motion. A series of ascending pitches may be co-perceived as a spatial rise or an acceleration, even when the tempo is unvarying. In short, an increase in

any of these features can be interpreted to issue from one or more of the others. They have found only small differences between trained and untrained subjects. Their findings lead them to suggest that there is interplay between musical “tension” and perceived content.

Their findings suggest that the analysis of isolatable musical features cannot reveal the dynamic processes at play in *listeners’* experiences. They pose the question Do intensity and the musical feature to which it is cross-attributed enhance one another? The answer appears to be a qualified yes. If there is a greater impression of one, is there necessarily less of the other? This is less clear.³ Eitan and Granot (2007) question why human listeners should respond to one changing attribute of music (e.g. increasing loudness) as though it were another (rising pitch). Their pairings cross several modalities: e.g. increasing tempo of a fix beat pattern as a crescendo; increasing dissonance in a series of chords as a crescendo. They also challenge the notion of “parametric additivity”—for example the concept of “joint-accent structure”, in which accentuation in one realm is held to strength coincident accentuation in another.

In their 2008 article the experimental material includes the intermingling of “true” and (at least partly) “false” compositions (inspired by Mozart) in order to evaluate whether music-theoretic ideas of thematic coherence (along the lines of Rudolf Reti’s postulates) can be verified. The short answer seems to be “no,” or at least “not convincingly so.” This study thus 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 as long as the grammar remains intact. In other words, semantic substitutions are quite tolerable provided they do not disturb the principles of the semantic architecture. Neither for trained not untrained subjects, listening to a hybrid work once or several times, did opinions vary significantly in their indifference to “coherence.” In the use of Mozart’s piano sonatas K. 280 and 332 as seed material, they found that in the case of the first vs. its hybrid, the majority of subjects preferred the hybrid. (In this regard, the 2008 study confirms central elements of Cope’s Emmy model.)

1E. Associative Similarity in Performance

Performance introduces its own experimental “noise” into our enquiry, because in the context of similarity studies it interjects several additional variables—among them timbre, mood, intensity, and tempo. Listeners perceive a whole. Few listeners are trained to distinguish one feature from another. This undifferentiated whole sets performed music apart from the short musical elements studied in controlled research environments. While in research contexts the pursuit of individual features (whether of the

³ In his *Expectation in Music* (2006), which delves deeply into an array of associative factors in music response, David Huron posits that many perceptions and reactions to musical phenomenon reflect animalistic interpretations of specific sound phenomena. The underlying relationships remain conjectural, because there is no accurate way to measure the totality of possible reactions to screeches, screams, hoots, and the like.

music itself or of its perception) is the norm, in ordinary listening one must process whatever is present quickly. To do that efficiently, though, it is likely that features are prioritized. What cannot be determined in present reckoning is whether such feature-priorities are stable for an entire piece, genre, repertory, or cultural context. In live performance, feature-interactions (and therefore the relative dominance of any single feature) have the power to produce continuing change in relative feature-rankings.

For the present we are left to fall back on anecdotal evidences. With respect to the kinds of features considered here, anecdotes from the artificial (computer-generated) music community are especially numerous. This abundance seems to owe primarily to the strength of opinion about computer-generated music. It is widely expected to seem “mechanical” in performance, and moral indignation is expressed when slight resemblance is noted between a known piece and an artificially composed one. Fortunately, there are two sources apart from the Haydn-Mozart Quartet Quiz (which involves only the variable of timbre) that are near to hand.

Artificial chorale harmonizations in live performance

The most extensive work involving the generation of new music on a well-known model from the eighteenth century co-existed with the first years of Emmy’s development. In a master’s thesis on expert systems Kemal Ebcioğlu (1986) reported the development of his efforts over several years to “discover” and implement--through extensive networks of voice-leading rules--satisfactory four-part harmonizations of Bach chorale melodies. This involved feeding the computer the chorale tune and generating three real parts. Although there had been (and continue to be) numerous research projects focused on harmonic analysis of encoded data, Ebcioğlu’s was the first sustained one devoted to generating new harmonizations.

To anyone who has not attempted such a fete, it may seem a trivial undertaking that could be largely accomplished by implemented the countless rules of harmonization that can be found in tutorials and manuals, but these rule systems invariably prove to be rudimentary. All procedures affecting compositional texture (harmony, counterpoint, voicing, orchestration) require a great deal of implicit knowledge that is nowhere codified. Ebcioğlu began with a few dozen rules derived from music-theory treatises but eventually turned to additional rules derived from observation and analysis. He continued to work on the system after completing the thesis. Eventually the rules numbered in the hundreds.

The hundreds of chorale harmonizations in which these rules were implemented showed consistent progress in defining note-based features. One thing that remained absent, however, was the implicit gestural knowledge that informs the works of human composers but, for lack of formulation, is difficult to legislate in computer-generated music. For this reason, many proved difficult to perform because of awkward hand requirements. Open- and closed-positions (the difference between a roughly even distribution of parts across the tonal range as opposed to a clustering of soprano, alto, and tenor at least an octave above the bass) were foreign to the computer. Although chorale harmonizations are short

works, the need to stretch one hand to a tenth and the other to a twelfth or thirteenth proved to be a common occurrence.

Apart from the lapse in gestural requirements that affects performers, the auditory impression of the generated music is more convincing when played live: users' perfunctory measure of "mechanical" music is inevitably a perception of metronomic precision, not compositional content. With human imprecision of the beat, the next trap for the listener seems to be timbre: if a mean-tone registration on a pipe organ is used to play an Ebcioğlu-Bach harmonization, almost no one in a live audience is likely to consider it computer-harmonized. Timbre is simply too dominant in its ranking in our perception of the musical details we are trained to perceive. Because of their brevity and overall simplicity of means, the computer-generated harmonizations that do not have obvious errors of voice-leading are unlikely to offend for lack of adherence to a template of musical style.

Artificial Compositions in Live Performance

Audiences appear to have consistent difficulty in distinguishing manmade from machine-made compositions if both are performed live. "Machine made" is, conversely, a judgment readily applied to works heard in quantized MIDI files irrespective of the origin of the content. In 1997 a concert in which computer-generated and man-made works were performed live was given at Stanford University. It formed part of a weekend symposium on human vs. machine creativity in music. A pianist and a violinist provided the music. No program was distributed until the end of the concert so that audience members could judge which ones were which on the basis of their own perceptions. Ten pieces (mainly concentrated in styles of the nineteenth century) were performed. Half were authentic, half were generated by Emmy (see Section 1A).

Although the idioms were familiar ones to an audience accustomed to classical music, none of the genuine works was particularly familiar. No piece was rated by more than 50% of the respondents as being authentic. No piece was rated lower than 10%. What the results strictly ruled out was Douglas Hofstadter's view (the fulcrum of the symposium) that without a genuinely "human" contribution to its composition, machine-composed music could not project a sense of human inspiration. (The extensive discussion, spread over many additional contributions and complemented by a large number of Emmy scores, is reproduced in Cope 2001.) Cope's view was that although a computer generated the work, all of the component parts from which its works were constructed originated in the human sphere.

2. Do Components of Associative Similarity mirror Baroque *Affetti*?

The Doctrine of Affects cultivated in the time of Monteverdi was not linked in any way to thoughts on musical similarity, which is a preoccupation of our own times. Yet there are some striking resonances between the musical motivations of certain associative correlates of similarity in recent studies with the musical motivations of *affetti* in the early seventeenth century. *Affetti* are usually discussed either in the context of Monteverdi's use of them or as a series of prescriptions for specific ornamental devises. There is

little reflection on their underlying purpose, which was to use the ear as a vehicle to leave an imprint on the mind. Certain kinds of *affetti* were more often than not used in the realm of sacred music than secular. Although it disapproved of “self-important” music, the Counter-Reformation considered “moving the hearer” to piety one of the noble goals of devotional music. In the absence of such a goal, musical ornamentation was merely a distraction from spiritual observance. (In other precincts during the Renaissance, emotional emblems were anticipated in general in the writings of Marsilio Ficino and Vincenzo Galilei. Sixteenth-century madrigals contained many harbingers of them. Collectively, *affetti* were to have many musical progeny in the elaboration of rhetorical concepts in music and in later German repertoires, particularly that of *Sturm und Drang*. Our discussion of *affetti* is confined to the Italian baroque.)

Despite their discussion in countless practical manuals of the time, we cannot know much about the actual execution of the *affetti*. There is a good chance, however, that the *stile concitato* was an instance of “intensity” along the lines pursued by Eitan-Granot. Although the formalization of the Doctrine of Affects is usually linked to the preface of the Book 8 of Monteverdi’s *Madrigali*, some notable works containing these affects were performed earlier.

At the level of expressing emotional turmoil in ways that were not yet crystalized, Monteverdi was already hard it in Mantua in 1607. The emotional import of his Sonata sopra “Sancta Maria ora pro nobis” (composed in 1607; published in 1610) was somewhat different from what it would later become: the obsessive dotted sixteenth-32nd pairs that appears in the work’s numerous instrumental parts are gradually calmed by the litany, so that with each utterance of it the number of notes per phrase declines. The dotted figure eventually disappears. The frenetic activity of the opening subsides into even half and eventually whole notes. There is much speculation as to why Monteverdi composed this unique piece.

The full-blown *stile concitato* made its grand entry (in the company of the *stile molle* and the *stile temperato*) in Monteverdi’s *Combattimento di Tancredi e Clorinda* in 1624, when the work was performed at the Palazzo Mocenigo in Venice. When it was finally published fourteen years later in the *Madrigali guerrieri ed amorosi* (Book VIII of Monteverdi’s madrigals), the composer sought to give the three “styles” (agitated, temperate, and calm) he was rigorously exploring in the fabric of ancient ideals—notably Plato’s belief that particular kinds of music had the power to affected human behavior in specific ways. The agitated style could make warriors brave; the calm style could render them morally weak and militarily defenseless; the temperate style suited periods of peace and security. This story from the capture of Jerusalem in the twelfth century was full of passion: Tancredi and Clorinda (who was disguised as a male warrior) were from opposing sides. The conflict between love and duty permeated every measure of the work. We cannot sequester the *stile concitato* steadfastly with bravery. It provided the context for jealousy (Example 1), wrath, frustration, anxiety, and the pressure of shortening time (another instance of intensity).

The image shows a musical score for a scene from Monteverdi's *Il combattimento di Tancredi e Clorinda*. It is divided into two systems. The first system features a vocal line (Testo) with the lyrics "A pas-si tar- di e len- ti Quai due to - ri ge -" and instrumental parts for Violino primo, Violino secondo, Viola, and Basso Continuo. The second system features a Soprano vocal line with the lyrics "lo - si di - r' ar - den - ti!" and instrumental parts for strings (S, T, B 1, B 2) playing a rhythmic pattern of eighth notes.

Example 1. Depiction of jealousy in Monteverdi's *Il combattimento di Tancredi e Clorinda* (composed 1624; published 1638).

In the wider world of Italian music over the first half of the seventeenth century, the *stile concitato* was much more widely employed than the other categories of affects. In many cases it rested on a long string of iterations of the same pitch in progressively shorter note values. In the all-embracing *stile rappresentativo*, in which early music dramas were conceived, each short idiom indulged the passion of the moment. Exaggeration was rampant. The elapsed time of text as metered by its musical setting was unpredictable, as a singer rushed through one word or phrase only to slow down on the next, always in the interest in highlighting the ever-changing emotional moment.

Among the works of the young instrumentalists who worked in Monteverdi’s Venetian orbit (and may well have performed in *Il combattimento*), were the ensemble sonatas of Dario Castello (from Venice), Biagio Marini (from Brescia), and Carlo Farina (from Mantua). Instantiations of the general prescription for use of the *stile concitato* varied. It seems to have been the foundation for the three predominant kinds of articulation used in instrumental music in the imaginative 1620s. These were the (measured) tremolo of wind instruments (and organs) and the two main ornaments of the early violin repertory—the groppo (corresponding in overall shape to the modern trill) and the trillo (corresponding to the long sequences of dotted notes, descending slowly but in short note values, in Monteverdi). For the measured tremolo, we find sterling examples in the canzonas of the recorder-player Giovanni Battista Riccio.⁴ The intended ensemble instruments in Example 2 are recorder (*flautino*) and bassoon (*fagotto*).

The image displays a musical score for three instruments: Recorder (Flautino), Bassoon (fagotto), and Basso continuo. The Recorder and Bassoon parts are marked with 'Tremolo' and feature rapid, repeated eighth-note patterns. The Basso continuo part consists of a few dotted notes. The score is presented in two systems, with the second system starting at measure 3.

Example 2. From the “Canzona ‘La Grimaneta’ with Tremolo” from Giovanni Battista Riccio’s *Terzo libro delle divine lodi* (1620/21).

⁴ The Counter-Reformation forbade the use of the recorder (which it claimed induced moral decadence) in sacred music, but Riccio’s instrumental pieces appeared in collections of sacred music. This suggests that ordinary parish churches ignored Papal injunctions. The Venetian Republic had had a big falling out with the Papacy in 1607 and was disinclined to follow every detail of proscriptions.

The tremolo passage shown in Example 2 was regarded as a marker of growing anxiety. Its cousin, the *trillo*, an incisive rather than a long-sustained expression, illustrated such things as a fluttering heart responding to momentous circumstances. For Marini (who titled his first collection of works, the *Affetti musicali* of 1617/8), there were several varieties of ornamental affects. Chief among them were the *grosso* (a melodic figure than metamorphosized into the modern trill) and a *trillo* (a crying figure similar to the one used in Monteverdi’s Sonata sopra ‘Sancta Maria ora pro nobis’. See Examples 3a and b.



Examples 3a and b. The *grosso* and the *trillo* figures used by Biagio Marini in his “Sonata con tre violini in eco” from Op. 8 (1626/9).

To summarize the implications of these examples, in *Tancredi* (Example 1), Monteverdi’s aim was to stir the listener into actually believing he too was at the gates of Jerusalem as the Crusaders sought to liberate the city. In modern audio-based studies of listeners, it is precisely in those hoped-for responses that Monteverdi’s aims seem to lead so nearly to the “user data” that industrial-strength data-mining algorithms seek to identify in casual listeners. The Baroque view of what the Counter-Reformation desired was listener *engagement*. Music was useless, in the view of the Council of Trent and its promoters, if its message was unheard and unheeded. Modern investigators may not be concerned with received messages, but they are very much concerned with listener engagement, for engagement is the threshold for memory and desire. Gino Stefani captured dozens of expressions of this view in the first volume of his *Musica barocca* (1974, 1987). In his view, the triple legacies of music as a discipline, as an art, and as a science were intricately combined in the seventeenth-century belief that music could move mountains if it could first engage the spirit of its listeners. Among all the arts, music was the one on the most intimate terms with Heaven, whose choirs inspired the affective singing of the *stile moderno*.

Style, mood, and intensity (1A, 1C, and 1E) can all be associated in one way or another with these affective categories. The *stile concitato* label pertained to homophonic passages notable for their rhythmic insistence but not for melodic or harmonic content. A series of running chords with little movement attracts the attention of listeners only insofar as its graduated rhythmic intensity or harmonic change engages the ear, as in Example 2, a short quotation of a *tremolo* passage in G. B. Riccio’s canzona “La Grimaneta” (1620-21). Since the work was scored for recorder and bassoon (with basso continuo), timbre (1B) alone would have engaged the audience, because these were still little used instruments in Italy.

It seems unlikely that listeners would *not* have noticed the difference between the two or between trombone and violin, another popular combination of the time).

Would listeners have been primed by the general discourse of the time to recognize and/or prioritize the sonic symbols of the *stile moderno*—*groppi* and *trilli* (Example 3) as well as *tremoli* and long passages of the *stile concitato*? The possibility represents a potential inversion (but not a negation) of the unexpected Aucouturier-Pachet audio pairings, which rest on general impressions of larger sound canvases. If there were some way to confirm the point, it would suggest that none of the current methods of audio-based clusters of works are disengaged from the rich legacies of the early Baroque.

Many musicologists and music theorists continue to regard music-oriented studies in algorithmic composition, perceptual psychology, sound engineering, audio retrieval, and kindred activities as foreign to their disciplines. Those steeped in Italian repertoires of Monteverdi's time are, however, well equipped to bring understanding to these lateral spheres of recent research. Gino Stefani's brilliant round-up of period descriptions to music of the seventeenth century shows a thousand times over that music had (in the collective mind of that time) no higher calling than to *move* its hearers. If it could not accomplish that, it was of no value to its sponsors (then the State, the Court, or the Church). The motives of today's Company (the largest sponsor of audio-based research) are far less noble, but unintended revelations may yet accrue. The search for shared principles of "similarity" may, in the context of musical associations, may ultimately reveal no more (or less) than the fundamental means by which music engages its hearers.

2011 Eleanor Selfridge-Field

References

- Aucouturier, Jean-Julien, and François Pachet (2002). "Finding Songs that Sound the Same," *Proceedings of the 1st IEEE Benelux Workshop on model-based processing and coding of audio (MPCA-2002)*, Leuven, Belgium.
- Aucouturier, Jean-Julien, and François Pachet (2008). "A Scale-free Distribution of False Positives for a Large Class of Audio Similarity Measures," *Pattern Recognition* 41/1, 272-284.
- Castello, Dario (1629). *Secondo libro di sonate concertato in stile moderno*. Venice.
- Cope, David (2001). *The Virtual Score: Computer Synthesis of Musical Style*. Cambridge, MA: The MIT Press.
- Cope, David (1991). *Computers and Musical Style*. Madison, WI: A-R Editions.
- Cope, David (1996). *Experiments in Musical Intelligence*. Madison, WI: A-R Editions.
- Cope, David (1998). "Signatures and Earmarks: Computer Recognition of Patterns in Music" in *Melodic Similarity, Concepts, Procedures, and Applications (Computing in Musicology, 15)*, ed. Walter B. Hewlett and Eleanor Selfridge-Field. Cambridge, MA: MIT Press.

- Cope, David (2001). *Virtual Music: Computer Synthesis of Musical Style*. Cambridge, Massachusetts: The MIT Press.
- Cope, David (2006). *Computer Models of Musical Creativity*. Cambridge, MA: 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 (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>.
- Marini, Biagio (1629). *Sonate, sinfonie, canzoni, passemuzzi, baletti, correnti, gagliarde, & retornelli : a 1, 2, 3, 4, 5 & 6 voci, per ogni sorte di instrumenti, un capriccio per sonar due violini quarttro parti, un eco per tre violini, & alcune sonate capricciose per sonar due e tre parti con il violino solo, con altre curiose & moderni invenzioni*. Book VIII. Venice.
- Monteverdi, Claudio (1638). *Madrigali guerrieri, et amorosi con alcuni opuscoli in genere rappresentativo, che saranno per brevi episodii fra in canti senza gesto*, Book VIII. Venice.
- Oakes, Steve, and Adrian C. North (2006). "The impact of background musical tempo and timbre congruity upon ad content recall and affective response," *Applied Cognitive Psychology* 20/4: 505-520.
- Patel, Aniruddh, L. R. Slevc, E. Federenko, D. Casasanto, J. Winawer, J. Rosenberg, and E. Gibson (2008). "Interference between syntactic processing in music and language," Cognitive Neuroscience Society Annual Meeting, San Francisco. Online: http://web.mit.edu/evelina9/www/Downloads/Posters/Patel_CNS2008_poster.pdf.
- Riccio, Giovanni Battista (1620/1). *Il terzo libro delle divine lodi*. Venice.
- Stefani, Gino (1974, 1987). *Musica barocca*, 2 vols. *I. Poetica e ideologia; II. Angeli e Sirene*. Milan: Studi Bompiani.
- Tzanetakis, George, Georg Essl, and Perry Cook (2001). "Automatic Musical Genre Classification of Audio Signals," International Symposium on Music Information Retrieval, Bloomington, IN. Online: ismir2003.ismir.net/papers/McKinney.PDF.