Exploring the variability of musical-emotional expression over historical time

JOSHUA FRANK [1]
Stellenbosch University and University of Cambridge

PETER M. C. HARRISON
University of Cambridge

BARRY ROSS
Independent Researcher

CARINA VENTER
Stellenbosch University

ABSTRACT: A listening experiment was designed to test whether modern listeners perceive the same affective content in Baroque music as the composer intended to portray. Listeners rated three musical examples from Johann David Heinichen's 1728 treatise *Der General-Bass in der Composition* for valence and arousal. Examples were chosen based on descriptions by the composer in which he outlined their intended affective content. Results showed a significant mismatch between original descriptions and listener ratings, which may indicate a change in the perceived affective content of the music. The historical variability of musical-emotional expression in general, with a focus on the role of structural emotion cues (particularly mode), is discussed, closing with suggestions for future research in the area of historical musical emotion.

EMPIRICAL research into musical emotions has been ongoing since the early 20th century. Kate Hevner (1935; 1936; 1937) presented the first rigorous studies examining multiple acoustic cues and their effects on the emotional interpretation of musical passages. The trends she identified for her limited set of acoustic cues predominantly align with the expectations of present-day Western performers, musicologists, and music theorists: the major mode carries associations with mostly positive emotional terms, the minor with negative (1935); firm rhythms are vigorous and dignified, flowing rhythms graceful (1936); fast tempo is exciting, slow tempo dreamy (1937); high pitch is playful, low pitch dignified (1937).

More recent studies have built upon Hevner's findings, largely operating within two theoretical paradigms: discrete emotion theory and dimensional theory. In the discrete paradigm, emotional effects are often identified over a somewhat flexible set of *basic emotions* (Scarantino & Griffiths, 2011). In the dimensional paradigm, emotional effects are identified using several *dimensions*, usually the two bipolar dimensions of valence (happy-sad or pleasant-unpleasant) and arousal (high energy-low energy or similar), after Russell (1980). These two paradigms have enabled a greater degree of comparison between studies, though definitional and methodological discrepancies remain prevalent.

Between and within paradigms, findings are inconsistent for many cues. However, some cues do show clear and consistent effects across studies of Western listeners. Mode stands out in this regard. Within the discrete emotion paradigm, the major mode is associated with happiness and tenderness, and the minor mode with anger, fear, sadness, and possibly disgust (Quinto *et al.*, 2014; Scherer & Oshinsky, 1977). In the dimensional paradigm, the major mode is associated with positive valence and the minor mode with negative valence, across various operationalizations (happy-sad bipolar scale: Gagnon & Peretz, 2003; Peretz *et al.*, 1998; pleasantness-unpleasantness bipolar scale: Scherer & Oshinsky, 1977; selection of positive vs negative facial expressions: Kastner & Crowder, 1990; Gregory *et al.*, 1996; Gerardi & Gerken, 1995).

The consistency in findings for the impact of mode on emotion judgments accords well with the experience of most Western listeners. However, it is somewhat surprising from an evolutionary perspective. While many of music's emotional cues can be evolutionarily rationalized in terms of patterns of vocal emotion expression (e.g., pitch level, spectral energy distribution, pitch contour, loudness; Juslin & Laukka, 2003), mode has no obvious analogue in non-musical vocalizations. Some researchers have nonetheless suggested candidate rationalizations for the emotional impact of mode, for example noting that melodies in the minor mode tend to consist of smaller intervals than those in the major mode, with these intervals potentially eliciting negative valence (Bowling *et al.*, 2012; Huron & Davis, 2012), or noting that the minor triad, which has more prominence in the minor mode, creates a greater degree of negatively valenced psychoacoustic roughness than the major triad (Helmholtz, 1863; Crowder, 1984). Cross-cultural research provides a mixed perspective on this question, showing evidence of both consistency and divergence in the valenced interpretation of roughness and mode across musical cultures (Fritz *et al.*, 2009; Athanasopoulos *et al.*, 2021; McDermott *et al.*, 2016).

An understanding of the emotional content of music in terms of specific structural elements and their relation to vocal expression (speech) is by no means a contemporary phenomenon. More than 400 years ago, at the turn of the 17th century, a group of Italian musical intellectuals, now commonly known as the Florentine Camerata, theorized the very same thing: that music may gain expressive properties through its imitation of the acoustic patterns of impassioned speech (Kivy, 1980). Vincenzo Galilei, resident theorist of the Camerata, recommended that musicians inform their approach to musical expression by listening to human speech in different emotionally laden contexts. He mentions speech rate, loudness, and the ascending or descending contour of the voice as elements worth noting. These ideas were echoed and implemented in the writings and compositions of musicians associated with the Camerata, such as Jacopo Peri and Giulio Caccini (LeCoat, 1972).

In 1618, René Descartes wrote his *Compendium Musicae*. While he only mentions one specific example of how structural features of music can relate to affect (associating faster tempo with "faster" emotions and slower tempo with "quieter" emotions; Descartes, 1618, see p. 15), his work became a cornerstone of later theory. Practicing music theorists such as Johann Mattheson expanded on Descartes' ideas, drawing also from his mid-century exposition of his theoretical psychology of emotion, *The Passions of the Soul* (Mattheson, 1739; Lang, 1967). For example, Mattheson states that wider intervals are associated with joy, and narrower intervals with sadness, based on the movements of the *esprits animaux* in these emotions, as detailed in *The Passions of the Soul*.

Mode, as a major/minor binary, was not as affectively clear-cut to Baroque theorists as interval size or tempo. Descartes stated that the major third and sixth were "more pleasing and more gay" than their minor counterparts (1618, p. 27); however, he declined to make explicit links between the (church) modes and emotions, though he acknowledged the division between modes with major and minor thirds in their "more prominent positions" (1618, p. 52). Mattheson, in *Das neu-eröffnete Orchestre* (1713), acknowledges the opinion that major modes are happy and minor modes sad but qualifies this as overly simplistic (Mattheson & Lenneberg, 1958). His presentation of characteristics for some of the 24 keys bears this out: most of the major keys are described as suitable for happier expressions and the minor keys for sadder ones, but there are exceptions. For example, E major and E-flat major are both attributed qualities of sadness, while D minor and G minor are said to be at least partially suited for happier music (Mattheson & Lenneberg, 1958). Mattheson also made clear his opinion that key characteristics are subjective rather than universal, noting that his descriptions are purely his own, not necessarily to be adopted by the reader of his treatise (Mattheson & Lenneberg, 1958).

Striking parallels exist between Baroque and modern empirical theories of musical emotional expression. Baroque prescriptions on the use of intervals and tempo are mirrored in modern music perception studies (e.g., Juslin & Laukka, 2003; Scherer & Oshinsky, 1977). However, due to the limited set of cues and emotions mentioned directly in the 17th- and 18th-century musical literature, it is not possible to make a confident claim from theory alone as to the similarity between Baroque and modern approaches to encoding (and decoding) emotional content in music.

One type of primary source that is optimally positioned for use in empirical investigation of this question is musical examples written by Baroque theorists and composers, explicitly to exemplify specific emotional content. Unfortunately, such examples are rare. The most promising set is to be found in Johann David Heinichen's *Einleitung* to his treatise on thoroughbass realization, *Der General-Bass in der Composition*, published in 1728 (Buelow, 1986). This, indeed, is the source to which Johann Mattheson directs the reader of *Der vollkommene Capellmeister* (1739, see p. 106). Heinichen provides a total of eighteen musical examples in his *Einleitung*, the majority of which are written to demonstrate the use of the *loci topici* as a route from text to affective composition (1728, see Appendix B).

In this paper, these musical examples are used to make a first foray into the experimental study of musical-emotional expression and perception across time within the Western art music tradition. The

question of *historical listening*, that is, how historical listeners perceived music, has been the subject of keen theoretical discussion (e.g., Burstyn, 1997), but experimental investigation has been sparse. We are aware of only one study (Stoessel *et al.*, 2021) comparing a historical musical phenomenon (a link between the semantic concept of "sweetness" and consonance in medieval music) with modern listeners' responses. To our knowledge, the present study is the first to compare listener ratings of emotion in historical music with the composer's explicit intention. Heinichen's musical examples will be used to address a question foundational to future study in this area: Do modern listeners hear the emotional content in Baroque music that the composer intended to portray?
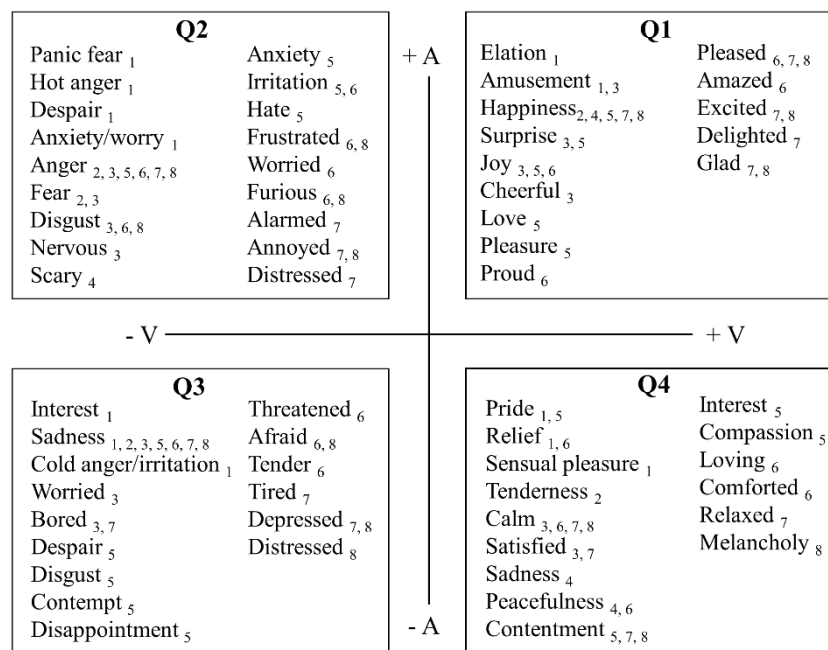
METHODS

Selection of Stimuli

Heinichen provides lengthy verbal descriptions for each of his musical examples. These are complex; in no case is a unitary emotion clearly and unambiguously designated by the text. In selecting examples to be used in the present experiment, subjective interpretation of Heinichen's descriptions was unavoidable but was based as far as possible on prior empirical literature. G. J. Buelow's English translation of Heinichen's treatise (Heinichen, 1728) was used as the source for all verbal content.

The subjective element of the selection procedure was the extraction of key emotion words from Heinichen's descriptions. All keywords were extracted while studying only the verbal descriptions (without reference to the sheet music). The procedure was to identify all emotion words included in each description, and to make a contextual evaluation of relative importance, selecting the most important word (one keyword per description). Appendix A contains Heinichen's descriptions for each excerpt, along with a demonstration of the emotion words selected. Where descriptions did not indicate affective content at all (some of the examples were written to demonstrate other rhetorical and compositional techniques), these and their corresponding musical examples were discarded. The keywords extracted from the nine eligible descriptions were *furious, rage, amorous, love* (two descriptions)*, flirtatious, tenderness, anxious,* and *playful*. Independent, though indirect, support for the appropriateness of these keywords is to be found in Mattheson (1739, see p. 106). Mattheson lists a set of affective labels by which he interprets Heinichen's examples. Included in Mattheson's list are the terms *rage, fear, play, love,* and *flirtatiousness*. These are either identical to or correspond closely with all the extracted keywords. Sadly, Mattheson does not describe the direct correspondences between these labels and specific examples. Keywords were extracted prior to reading this passage in Mattheson. An important note to make here is that the selection of keywords and corresponding excerpts as representative of actual listener perceptions in the 18th century, even in localized form, relies on the idea that Heinichen was, in his own time, successful in his musical-emotional portrayals. This assumption cannot be validated empirically and is made solely based on Heinichen's positive reputation in his day and Mattheson's choice of Heinichen's examples as illustration of his own (popular) theories. This caveat should be borne in mind when interpreting the results below.

Next, the extracted keywords were mapped to quadrants of the valence-arousal (VA) space. There are four such quadrants: Q1 (high V, high A), Q2 (low V, high A), Q3 (low V, low A), and Q4 (high V, low A). In principle, one could adopt a finer-grained mapping between keywords and VA space, but we reasoned that our relatively coarse-grained approach would be more robust to uncertainties in word translation and changes in word usage over time. Keywords were categorized with reference to two previous music perception studies in which emotion terms were mapped to dimensional space (Eerola & Vuoskoski, 2011; Vieillard *et al.*, 2008) and similar studies from non-musical domains (Mehu & Scherer, 2015; Hupont *et al.*, 2013; Fontaine *et al.*, 2007; Richins, 1997; Russell, 1980; Morgan & Heise, 1988), producing the categorizations described in Figure 1. Not all keywords were present in previous work; in these cases, the words were categorized manually with guidance from the emotion taxonomy of Shaver *et al.* (1987). This source was also referred to in cases where a keyword or closely related word appeared with equal frequency in more than one quadrant in the reference literature. While this approach to categorization would certainly be entirely alien to Heinichen and other Baroque musicians, it allows for easy data gathering and comparison. Approaches utilizing discrete emotion categories or other methods could also be used.

In the next step, representative musical examples were selected for different quadrants of the VA space. Unfortunately, Heinichen's descriptions provided no good examples for quadrant 3 (low arousal, low valence). For each of the remaining quadrants, one example was selected from Heinichen, alongside one example from a set of modern excerpts (Vieillard *et al.*, 2008) for which VA ratings have already been experimentally established. The inclusion of the modern excerpts allows for validation of the experimental procedure by comparison of listener ratings in the present experiment with the ratings

established by Vieillard *et al.* (2008). In total, six examples were selected: two paired examples, one Baroque and one modern, for each of the three eligible quadrants. It should be noted that this is a very small stimulus set, representing the work of only a single composer for the Baroque excerpts; suggestions of how to increase the number and diversity of stimuli in future experiments, to obtain more reliable results, are made toward the end of this article. The selected examples from Heinichen corresponded to the keywords *playful* (quadrant 1), *furious* (quadrant 2), and *love* (quadrant 4); these examples will hereafter be referred to as E1, E2, and E4. These were paired with Vieillard *et al.*'s examples G03, P07, and A07, which prior empirical ratings had placed in quadrants 1, 2, and 4, respectively (Vieillard *et al.*, 2008). All examples were shortened to end on a chord with pre-dominant function, to avoid inducing closure-related feelings that could confound the evaluation of the primary emotion of each stimulus. Additionally, paired examples were constructed to differ in length by no more than two seconds. Appendix A provides the examples in full, as used in the experiment.

Q2			Q1	
Panic fear $_1$	Anxiety $_5$	+ A	Elation $_1$	Pleased $_{6,7,8}$
Hot anger $_1$	Irritation $_{5,6}$		Amusement $_{1,3}$	Amazed $_6$
Despair $_1$	Hate $_5$		Happiness $_{2,4,5,7,8}$	Excited $_{7,8}$
Anxiety/worry $_1$	Frustrated $_{6,8}$		Surprise $_{3,5}$	Delighted $_7$
Anger $_{2,3,5,6,7,8}$	Worried $_6$		Joy $_{3,5,6}$	Glad $_{7,8}$
Fear $_{2,3}$	Furious $_{6,8}$		Cheerful $_3$	
Disgust $_{3,6,8}$	Alarmed $_7$		Love $_5$	
Nervous $_3$	Annoyed $_{7,8}$		Pleasure $_5$	
Scary $_4$	Distressed $_7$		Proud $_6$	

− V ———————————————————————— + V

Q3		Q4	
Interest $_1$	Threatened $_6$	Pride $_{1,5}$	Interest $_5$
Sadness $_{1,2,3,5,6,7,8}$	Afraid $_{6,8}$	Relief $_{1,6}$	Compassion $_5$
Cold anger/irritation $_1$	Tender $_6$	Sensual pleasure $_1$	Loving $_6$
Worried $_3$	Tired $_7$	Tenderness $_2$	Comforted $_6$
Bored $_{3,7}$	Depressed $_{7,8}$	Calm $_{3,6,7,8}$	Relaxed $_7$
Despair $_5$	Distressed $_8$	Satisfied $_{3,7}$	Melancholy $_8$
Disgust $_5$		Sadness $_4$	
Contempt $_5$		Peacefulness $_{4,6}$	
Disappointment $_5$	− A	Contentment $_{5,7,8}$	

Figure 1. Mapping of selected emotion words to the VA space. Subscript numbers refer to sources. 1: Mehu & Scherer, 2015. 2: Eerola & Vuoskoski, 2011. 3: Hupont *et al.*, 2013. 4: Vieillard *et al.*, 2008. 5: Fontaine *et al.*, 2007. 6: Richins, 1997. 7: Russell, 1980. 8: Morgan & Heise, 1988.

Participants

Participants were all undergraduate students at Stellenbosch University ($N = 30$, 14 females, 16 males). Participants' mean age was 22 ($SD = 7.14$), ranging from 18 to 57. A wide range of practical musical experience and formal training was represented, ranging from no experience whatsoever to music students with up to 15 years of formal training ($M = 4.34$ years of formal training, $SD = 5.02$). No participants were unfamiliar with Western music, but familiarity with Baroque music was not assessed. Ethical clearance was obtained from the Stellenbosch University Research Ethics Committee: Humanities. Permission to gather data from students was obtained from the Stellenbosch University Division for Information Governance.

Playback

Musical examples were played back to participants as .WAV files, exported in 32-bit quality from Sibelius, version 8.7. All examples were played back using the "piano" timbre, in equal temperament, at a pitch level of A = 440Hz. Timbre and temperament could well be relevant to emotion perception in these excerpts, but lacking specifications of these parameters from Heinichen, we defaulted to the options that were simplest to implement, which are also likely the most familiar to participants. Stylistically appropriate tempi were suggested by a local Baroque music expert (who was not exposed to the verbal descriptions). A Dell Latitude 7280 laptop computer and Sennheiser HD 419 over-ear headphones were

used for playback. The volume was set to a comfortable level and held constant across experimental sessions.

Procedure

All experimental sessions were conducted one-on-one. Participants first filled out a questionnaire to capture demographic and other information. They were then given a copy of the rating sheet, and the experimental process was explained to them. Ratings were gathered using an 11-point scale for each dimension, ranging from -5 to 5. An 11-point scale was chosen for its greater degree of specificity compared to scales with fewer points, as well as the presumed ease of interpretation of five points on either side of a neutral midpoint. The scales were anchored verbally at their poles with "sad" and "happy" for valence, and "sleepy" and "energetic" for arousal. The zero point was labeled "neutral" in both scales. Each example was played twice. The participants filled in a valence score for the example on the first listening, and an arousal score on the second. Participants rated all six examples, which were presented in one of four pre-determined randomized orders.

Data Analysis [2]

All data analysis was done in *R* version 4.0.5 (R Core Team, 2021), making use of the *FSA* package (Ogle *et al.*, 2021). Visual representations were created in *R* using the *ggplot2* package (Wickham, 2016).

Cue Extraction

Numerical levels of three cues were manually extracted from Heinichen's examples for use in explanation of the relationship between listeners' ratings and the ratings expected from Heinichen's descriptions (see Figure 4 below). The cues chosen are rate of event presentation (REP, measured as onsets per second), mean pitch height (measured in semitones from middle C), and mean melodic interval size (measured in semitones). Pitch height was measured without regard to note length. Interval size was calculated within voices and then summed, taking the larger interval size where ambiguous, and was not calculated over rests. Additionally, mode was examined as a binary variable (major/minor). These cues were chosen as they are relatively easy to calculate by hand and have all been firmly linked to systematic effects on perceived valence and arousal in music (as discussed below).

RESULTS AND DISCUSSION

Valence and arousal ratings for all examples are shown in Figure 3. The three modern excerpts' ratings placed them in the expected quadrants. For the modern representatives of Q1 (median V = 4, median A = 3.5) and Q2 (median V = -1, median A = 1.5), this placement was unambiguous. For the representative of Q4 (median V = 0, median A = -2), the valence rating placed it on the border between Q4 and Q3; however, taken numerically, ratings of valence were slightly more positive than negative (mean = 0.13). Overall, these results echo those of Vieillard *et al.* (2008) and validate the present procedure.

Kruskal-Wallis tests indicated that there were significantly different valence ratings ($H(5)$ = 105.99, $p < .005$) and arousal ratings ($H(5)$ = 101.7, $p < .005$) among the six examples, as would be expected from musical excerpts chosen to illustrate different emotions. However, on closer inspection, the patterns of ratings diverge significantly from the patterns implied by Heinichen's descriptions. In particular, one may expect E2 (*furious*) to have the lowest valence of the three examples, yet we find instead that participants' valence ratings for E2 (*furious*) were not significantly different (median = 4) to those for E4 (*love*) (median = 4, $p = .89$, post-hoc Dunn's test with Bonferroni correction), and significantly *more* positive than those for E1 (*playful*) (median = 0, $p < .005$). Furthermore, one may expect E4 (*love*) to have the lowest arousal of the three examples, yet we find instead that participants' arousal ratings for E4 (*love*) were not significantly different (median = 3) to those for E2 (*furious*) (median = 4, $p = .74$), and significantly *higher* than those for E1 (*playful*) (median = 1.5, $p < .005$).

In addition to the relationships between Heinichen's examples reported above, all differences in valence and arousal between paired examples (Baroque and modern examples expected to fall into the same quadrants) were significant (all $p < .05$). For all pairs, differences in arousal values suggested that Heinichen's descriptions were not matched by listener ratings; that is, all arousal values for Baroque examples were significantly lower than their modern counterparts when Heinichen's descriptions implied high arousal, and significantly higher where low arousal was implied. The same was true of valence ratings for the paired representatives of Q1 and Q2. However, for Q4, E4 (*love*) had a significantly higher valence rating than its modern counterpart. Thus, E1 (*playful*) and E2 (*furious*) were significantly less

representative of their expected quadrants than their modern counterparts for both valence and arousal, while E4 was more representative in terms of valence and less in terms of arousal.

While the relationships *between* examples as captured in terms of valence and arousal, especially between Heinichen's examples, are more important than their exact placements on the VA space (as no historical baseline for valence and arousal ratings of different emotions can be established), it is worth noting that none of the Baroque examples fell into the quadrants best corresponding to their descriptions. E2 (*furious*) and E4 (*love*), for which Heinichen's descriptions best fit Q2 and Q4 respectively, were both rated clearly in Q1. E1 (*playful*), theoretically representative of Q1, is more ambiguous, with an appropriate (positive) median arousal rating, but an approximately zero median valence rating.

(a)   **(b)**
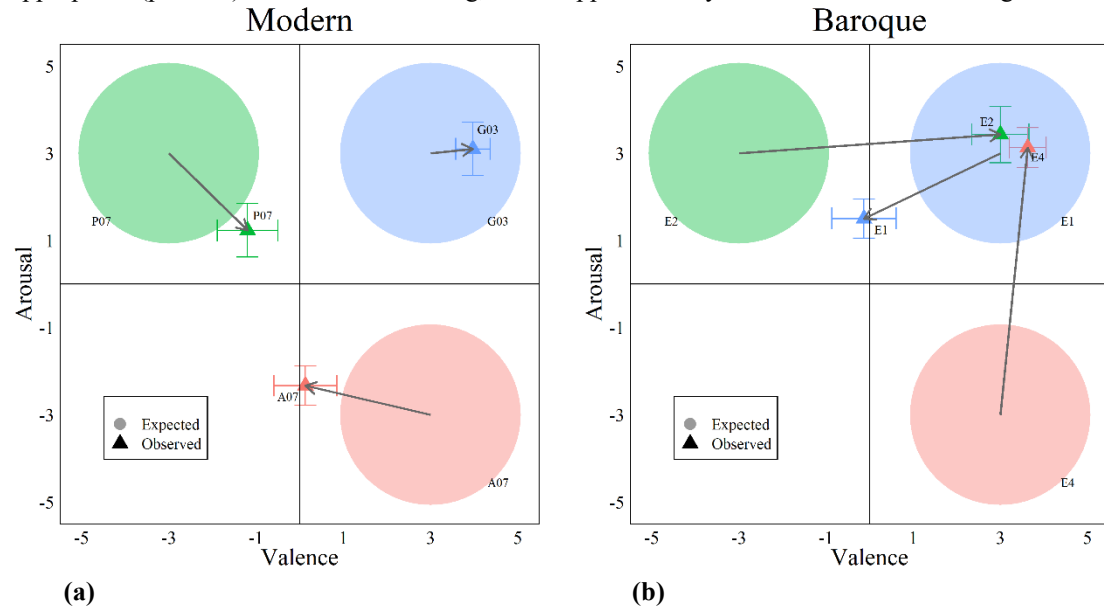
Figure 2. Scatter plot of observed vs expected mean ratings (± standard errors for valence and arousal) for modern (a) and Baroque (b) examples. Large circles, labelled "Expected", refer to rough placements expected for each of the examples. For Heinichen's excerpts, these expectations were based on the present interpretation of written descriptions accompanying these excerpts; for modern excerpts, they were based on the quadrant ratings found for these excerpts by Vieillard *et al.* (2008). Note the sizable discrepancy between expected and observed ratings for the Baroque examples (no means fell in the expected quadrant).
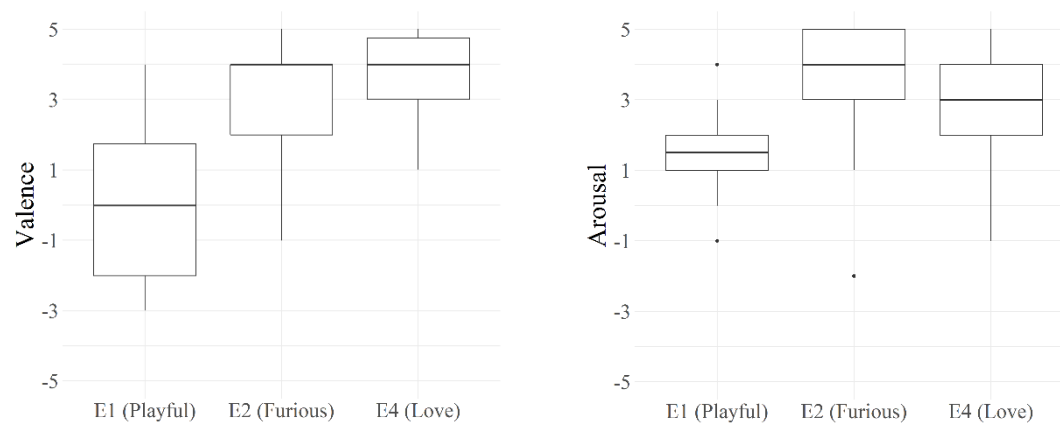
Figure 3a. Box plots of valence and arousal ratings for Heinichen's examples.
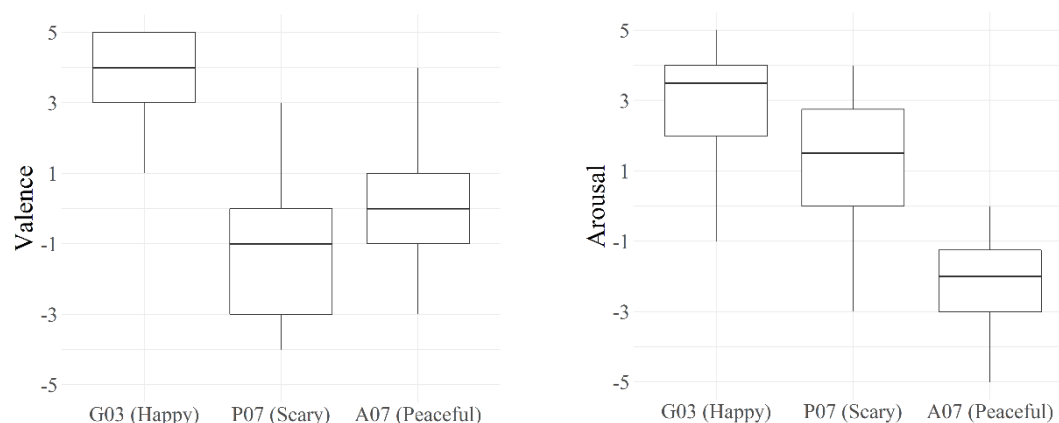
Figure 3b. Box plots of valence and arousal ratings for the modern examples.

The results of the present experiment suggest that modern-day listeners interpret Heinichen's examples differently from the composer's intentions, with the caveat that these intentions may have been distorted by translating complex verbal descriptions into VA space. Along the dimensions of valence and arousal, listener ratings differed significantly from the patterns suggested by Heinichen's descriptions of his examples. The quadrant-level mismatch is indicated in Figure 2b, but even by the more conservative metric of relationships *between Heinichen's excerpts* in VA space, the observed ratings do not align with the historical descriptions. Heinichen's descriptions implied that E2 (*furious*) would receive lower valence ratings and that E4 (*love*) would receive lower arousal ratings; modern listeners in the present experiment rated E1 (*playful*) the lowest in both valence and arousal. These results may be indicative of systematic changes in listeners' perception of musical emotions over time.

Shifting patterns of relationships between musical-structural emotion cues (Horn & Huron, 2015) and changes in average cue levels (Daniele & Patel, 2013) over time have previously been demonstrated in large corpus analyses of Western music. Changes in cue use over large corpora may reflect the changing prominence of different emotion portrayals (as concluded by Horn & Huron, 2015), but could additionally hint towards changes in the emotional impact of the cues. If modern listeners' interpretations of Baroque music differ from the composer's intentions, as the present results may indicate, this would support the latter interpretation.

Modern performers would be hesitant to claim knowledge of the precise compositional intention behind the music they interpret, but a feeling that one's interpretation is "appropriate", at least in broad terms, is integral to confident creative decision-making. Should the results of this study be generalizable, such "appropriateness" may become more difficult to assert, at least on a philosophical level. On a different note, alongside contemporary cross-cultural studies, understanding historical variability in the emotional "content" of music (as encoded by composers and performers and perceived by listeners) could help differentiate between the universal and the culturally variable in musical emotion. Exploring the stability or instability of emotional portrayals in music from other eras of Western art music history would be a fascinating project. However, the rest of this discussion will retain the focus on Baroque music, and further investigate the mismatch in emotional perception observed in the present experiment.

It is possible to interpret the present results as indicating that the use of structural cues in the communication of musical emotion has changed over time. If this is the case, the question of which cues are responsible for this change must be addressed. The structural element that most immediately presents itself as an explanation for the present results is mode. As discussed above, mode has been shown to have strong effects on valence ratings. The setting of E1 (*playful*) in the minor mode is unexpected from a modern perspective, as the positive end of the valence dimension is usually associated with the major mode. Likewise, E2 (*furious*) being in the major mode is unexpected, as this affect falls at the negative end of the valence dimension. These modalities neatly correspond with the mismatched valence ratings of E1 and E2 in the present experiment. Assuming that a typical listener in Heinichen's time would indeed have interpreted E1 as playful and E2 as furious, it follows that the historical impact of mode on emotion judgment may well have been different from that observed in Western listeners today.

A finding that emotional connotations of modes have changed over time speaks against biological theories of modes' emotional impact (e.g., Helmholtz, 1863; Bowling *et al.*, 2012). Such theories would demand that the major/minor binary (where it exists) be universally valenced, implying a historically static interpretation of mode. The lack of a clear affective binary concerning the major and

minor modes in the theoretical writings of other Baroque authors, as discussed above, further undermines such theories. It should be remembered, however, that the major/minor binary, though certainly present, did not dominate Western music in the period under consideration to the same degree as it does today. If differences in interval sizes between melodies in the two modes are partly responsible for their emotional impact, this may have manifested less strongly due to the relative abundance of older music composed in the church modes (i.e., due to the weaker representation of the major/minor binary in the music people heard).

Heinichen's examples survive only in notation; thus, only cues that can be (and have been) encoded in this medium are available for analysis. Many cues which could in theory be encoded in a musical score do not appear in Heinichen's examples. In the 17th and 18th centuries, it was common practice to leave a great degree of control over the sonic product in the hands of the performer. Heinichen does not systematically indicate, for example, dynamics, articulation, or even instrumental timbre: the melody instrument is not specified for many of his examples. The scope of cues available for analysis is therefore rather limited. However, future research could investigate other cues that are available for extraction from the scores, e.g., rhythmic and harmonic complexity.
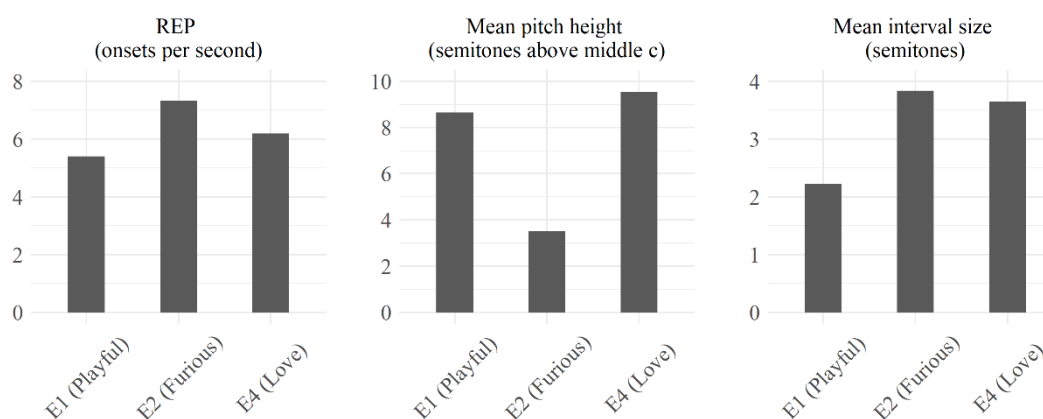
Figure 4. Levels of cues for Heinichen's examples. E1 (*playful*) is set in the minor mode, while E2 (*furious*) and E4 (*love*) are in the major.

Tempo (as REP), interval size, and pitch height are treated as analogous to speech elements far more commonly than mode (e.g., Juslin & Laukka, 2003; Bowling *et al.*, 2012; Huron & Davis, 2012). Having seen that Heinichen does not use mode in the way contemporary Western listeners tend to expect, it is worth assessing whether this is true of his use of other cues (i.e., whether the use of these cues has changed over time or remained invariant).

Tempo, alongside mode, is another cue that shows remarkable consistency in its effects across studies. It is associated positively with both valence and arousal (Scherer & Oshinsky, 1977; Schubert, 2004; Gagnon & Peretz, 2003; Ilie & Thompson, 2006). REP levels in the excerpts do not fit expectations based on Heinichen's descriptions: E1 (*playful*) may be expected to have a high REP, but has the lowest. This mismatch could indicate a shift in the affective connotations of tempo over time. It is worth noting that REP is influenced by BPM (beats per minute), and BPM levels were provided by a modern expert rather than by Heinichen himself, undermining the certainty of the mismatch.

Pitch height and interval size have both been positively correlated with arousal; interval size has also been positively correlated with valence, while pitch height has been negatively correlated with valence (Scherer & Oshinsky, 1977; Ilie & Thompson, 2006). E4 (*love*) has the highest mean pitch of the three examples, which does not match expectations based on Heinichen's descriptions. E2 (*furious*) has the largest mean interval size, which matches expectations in terms of arousal, but not valence; this reveals an issue of ambiguity that arises within the dimensional paradigm when a cue level has associations with both valence and arousal. Based on Heinichen's descriptions, E1 (*playful*) may be expected to have a larger mean interval size than observed, and E2 (*furious*) may be expected to have a higher mean pitch. It should be noted that these expectations are made based on the VA mappings of the emotion keywords for each excerpt, rather than the emotion terms themselves, and alternative operationalizations of these keywords might well lead to different expectations.

Heinichen's descriptions and a cursory analysis of cue levels seem to indicate that mode, pitch height, and interval size have shifted in their emotional connotations over time. Tempo shows the same trend, but this must be interpreted with additional caution due to BPM not being notated in the score. The shifting impact of mode may be explained by its lack of a direct analog in speech. However, pitch

height, interval size, and tempo are cues shared with speech, which may imply a biological basis for their ability to convey specific emotions. This would lead to the expectation that they would remain static over time, which is not borne out here.

The results of this study point towards the possibility that musical emotion perception in Western listeners has changed over the past 300 years, with changes in the emotional impact of structural cues being a candidate explanation for this effect. However, there are several important limitations of this work that must be noted before drawing firmer conclusions. In particular, the number of musical examples used in the experiment was extremely small, and all examples were drawn from a single composer. A possible approach for further investigations in this area would be to draw stimuli from lesser-known Baroque vocal music, using the text and context as indicators of intended emotional content. While Heinichen's examples were ideal in their relative obscurity, participants' familiarity with other stimuli could easily be captured in a questionnaire. It is also worth noting that Heinichen's examples do not exactly conform to Baroque theories, leaving ambiguity as to whether they may be considered representative at all, a problem that further studies using more diverse stimuli may help to resolve. Future studies could also measure levels of a greater number of cues and compare these to listener ratings to further explore and substantiate the idea of cues changing in their emotional impact over time.

In the above discussion, the efficacy of Heinichen's examples in his own time is taken as given, i.e., it is assumed that a typical Baroque listener, or at least one from Heinichen's own time and place, would interpret the examples as the composer intended. This assumption is made based on Heinichen's stature as a composer and theorist in his day, as well as Mattheson's endorsement of his examples, but is not supported by any direct evidence. It remains possible, therefore, that these examples did not function as intended in their own time, which would undermine the results of this study. Again, the use of a greater number of more diverse stimuli in future experiments would help in navigating this issue, as common patterns of cue use may be found across composers; these would then be assumed to reflect successful techniques of emotion portrayal.

A further limitation comes from our approach of reducing Heinichen's complex descriptions into single emotion keywords which were then mapped to the VA space. An alternative approach would be to avoid keywords altogether, and instead instruct participants to rate the verbal texts directly for valence and arousal. This approach could avoid some oversimplifications potentially stemming from the keyword approach: for example, Heinichen's description for E4 was narrowed down to the word *love* but may suggest a more energetic species of love — perhaps one to which both a modern and Baroque listener would ascribe higher arousal, as was seen in the present results. Likewise, E1 could potentially be categorized as *love* rather than *playful*, though its low valence rating relative to the other Baroque examples would still indicate a mismatch in this case. A further alternative would be to utilize a pairing paradigm, with listeners matching musical excerpts with the text directly; this would avoid both keyword and VA-translation simplifications of the text and may be the approach that best preserves the integrity of the verbal material, especially where this material is complex.

A follow-up study could investigate Heinichen's examples further. In the present experiment, only small segments of the examples were used (from the opening/introductory section of each). The intended emotional content of the examples may be better embodied in other sections. It would be interesting to compare the data from the present experiment with data generated when more of each example is heard. Such an experiment could make use of a greater number of examples from Heinichen and would be profitably accompanied by empirical rating of the textual descriptions, as discussed above. The explicit descriptions offered by Heinichen also present a unique opportunity for empirical investigation in the field of historically informed performance practice (HIPP). If patterns of cue use have indeed changed so significantly that modern listeners cannot decode the intentions of Baroque composers, perhaps HIPP performers could act as intermediaries, translating between cue uses. HIPP approaches to both the interpretation of a composition's intended affect and the portrayal of this affect in performance are to an extent informed by historical information, which has the potential to increase communicative efficacy at the level of composer-to-performer, but perhaps also to decrease it at the level of performer-to-audience. Both levels could be investigated by acquiring VA ratings of text and music from HIPP performers and listeners in the context of a HIPP performance. Such an experiment could be of great value to HIPP performers wishing to streamline this particular aspect of musical communication and could also present empirical support for the HIPP approach in the face of modern detractors.

The question of temperament is a perennial one in research on Baroque music. In Heinichen's time, a variety of temperaments could be heard in use, and preferences differed markedly. In the present experiment, equal temperament was used, primarily out of convenience. Further experiments could present musical examples in various historical temperaments and investigate the impact of temperament on listener ratings. It is worth pointing out that temperament could well play a role in the efficacy of the major/minor binary as a vehicle for carrying clear expressive connotations. In physical and

psychoacoustic terms, a true binary is only present between the two modes in equal temperament. In historical irregular temperaments, common in the 18th century, individual keys had unique combinations of interval sizes, making them sound different. This no doubt reduced the perceptual salience of the major/minor binary and may explain why the connotations of the modes as established in modern empirical research (which works almost exclusively with equal temperament) seem not to appear systematically in Heinichen's descriptions. Baroque theorists were more inclined, perhaps because of the temperaments they heard in use, to designate affective content at the level of key, rather than mode.

A final point worth exploring in future research is whether the mismatch between composer intentions and listener ratings may be reduced by presenting the music in a more ecologically valid manner. In the present experiment, MIDI playback with a single instrumental timbre was used for the presentation of the examples, and performance cues such as dynamics, timing, and articulation were not varied. Instrumental timbre and performance cues have been found to play a sizeable role in perceptions of musical emotion (e.g., Quinto *et al.*, 2014; Juslin & Laukka, 2003; Balkwill & Thompson, 1999; Juslin, 2000). The pursuit of more ecologically valid stimuli is worthwhile, though it presents non-trivial difficulties regarding the choice of interpretive approach. The question of performance cues could be investigated as part of the HIPP experiment mentioned above.

The present experiment has found empirical support for the historical variability of the affective impact of Baroque musical compositions. Findings of historical variability in the perception of this music are important not only in the purely academic sphere, but also in the practical field of Baroque performance practice. The results of this experiment hint towards an intriguing and difficult question, which future research in this field will refine and shape over time. That question is: For historically informed performers of Baroque music, ought the goal be to pursue the historical *sounds* of the music, or its historical *effects*?

ACKNOWLEDGEMENTS

NOTES

[1] Correspondence can be addressed to: Joshua Frank, Centre for Music and Science, Faculty of Music, University of Cambridge, 11 West Rd, Cambridge CB3 9DP, United Kingdom, jbf43@cam.ac.uk.

[2] Data are available on request from the first author.

REFERENCES

Athanasopoulos, G., Eerola, T., Lahdelma, I., & Kaliakatsos-Papakostas, M. (2021). Harmonic organisation conveys both universal and culture-specific cues for emotional expression in music. *PLoS ONE, 16*(1), e0244964. https://doi.org/10.1371/journal.pone.0244964

Balkwill, L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception: An Interdisciplinary Journal, 17*(1), 43-64. https://doi.org/10.2307/40285811

Bowling, D. L., Sundararajan, J., Han, S., & Purves, D. (2012). Expression of emotion in Eastern and Western music mirrors vocalization. *PLoS ONE, 7*(3), e31942. https://doi.org/10.1371/journal.pone.0031942

Buelow, G. J. (1986). Heinichen's *Einleitung* to the *General-Bass* Treatise: A translation. In Heinichen, J. D. (1728). *Der General-Bass in der Composition*. (G. J. Buelow Ed., Trans.). *Thoroughbass Accompaniment according to Johann David Heinichen* (Revised Ed.) (pp. 307-308). Ann Arbor, Michigan: UMI Research Press.

Burstyn, S. (1997). In quest of the period ear. *Early Music, 25*(4), 692-701. https://doi.org/10.1093/em/25.4.692

Crowder, R. G. (1984). Perceptions of the major/minor distinction: I. Historical and theoretical foundations. *Psychomusicology, 4*(1-2), 3-12. https://doi.org/10.1037/h0094207

Daniele, J. R., & Patel, A. D. (2013). An empirical study of historical patterns in musical rhythm: Analysis of German & Italian classical music ssing the nPVI equation. *Music Perception: An Interdisciplinary Journal, 31*(1), 10-18. https://doi.org/10.1525/mp.2013.31.1.10

Descartes, R. (1618). *Compendium Musicae*. (W. Robert Trans). *Compendium of Music*. American Institute of Musicology, 1961.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*(1), 18-49. https://doi.org/10.1177/0305735610362821

Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18*(12), 1050-1057. https://doi.org/10.1111/j.1467-9280.2007.02024.x

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology, 19*, 573-576. https://doi.org/10.1016/j.cub.2009.02.058

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cognition and Emotion, 17*(1), 25-40. https://doi.org/10.1080/02699930302279

Gerardi, G. M., & Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception: An Interdisciplinary Journal, 12*(3), 279-290. https://doi.org/10.2307/40286184

Gregory, A. H., Worrall, L., & Sarge, A. (1996). The development of emotional responses to music in young children. *Motivation and Emotion, 20*(4), 341-348. https://doi.org/10.1007/BF02856522

Heinichen, J. D. (1728). *Der General-Bass in der Composition*. (G. J. Buelow Ed., Trans.). *Thoroughbass Accompaniment according to Johann David Heinichen* (Revised Ed.). Ann Arbor, Michigan: UMI Research Press, 1986.

Helmholtz, H. L. F. von. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. (A. J. Ellis Trans.). *On The Sensations of Tone as a Physiological Basis for the Theory of Music* (4th Ed.). London: Longmans, Green, and co., 1912.

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology, 47*(1), 103-118. https://doi.org/10.2307/1416710

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

Hevner, K. (1937). The affective value of pitch and tempo in music. *The American Journal of Psychology, 49*(4), 621-630. https://doi.org/10.2307/1416385

Horn, K., & Huron, D. (2015). On the changing use of the major and minor Modes 1750–1900. *Music Theory Online, 21*(1). https://doi.org/10.30535/mto.21.1.4

Hupont, I., Baldassari, S., & Cerezo, E. (2013). Facial emotion classification: from a discrete perspective to a continuous emotional space. *Pattern Analysis and Applications, 16*(1), 41-54. https://doi.org/10.1007/s10044-012-0286-6

Huron, D., & Davis, M. J. (2012). The harmonic minor scale provides an optimum way of reducing average melodic interval size, consistent with sad affect cues. *Empirical Musicology Review, 7*(3-4), 103-117. https://doi.org/10.18061/emr.v7i3-4.3732

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception: An Interdisciplinary Journal, 23*(4), 319-330. https://doi.org/10.1525/mp.2006.23.4.319

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance, 26*(6), 1797-1813. https://doi.org/10.1037//0096-1523.26.6.1797

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770-814. https://doi.org/10.1037/0033-2909.129.5.770

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception: An Interdisciplinary Journal, 8*(2), 189-201. https://doi.org/10.2307/40285496

Kivy, P. (1980). *The corded shell: Reflections on musical expression*. Princeton, New Jersey: Princeton University Press.

Lang, P. H. (1967). The enlightenment and music. *Eighteenth-Century Studies, 1*(1), 93-108. https://doi.org/10.2307/3031668

LeCoat, G. G. (1972). Comparative aspects of the theory of expression in the Baroque age. *Eighteenth-Century Studies, 5*(2), 207-223. https://doi.org/10.2307/2737918

Mattheson, J. (1739). *Der vollkommene Capellmeister*. (E. C. Harriss Ed., Trans.). *Johann Mattheson's* Der vollkommene Capellmeister*: A Revised Translation with Critical Commentary*. Ann Arbor, Michigan: UMI Research Press, 1981.

Mattheson, J., & Lenneberg, H. (1958). Johann Mattheson on affect and rhetoric in music (II). *Journal of Music Theory, 2*(2), 193-236. https://doi.org/10.2307/843199

McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature, 535*, 547-550. https://doi.org/10.1038/nature18635

Mehu, M., & Scherer, K. R. (2015). Emotion categories and dimensions in the facial communication of affect: An integrated approach. *Emotion, 15*(6), 798-811. https://doi.org/10.1037/a0039416

Morgan, R. L., & Heise, D. (1988). Structure of emotions. *Social Psychology Quarterly, 51*(1), 19-31. https://doi.org/10.2307/2786981

Ogle, D. H., Wheeler, P., & Dinno, A. (2021). *FSA: Fisheries Stock Analysis*. R package version 0.8.32, https://github.com/droglenc/FSA.

Peretz, I., Gagnon, L, & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition, 68*, 111-141. https://doi.org/10.1016/S0010-0277(98)00043-2

Quinto, L., Thompson, W. F., & Taylor, A. (2014). The contributions of compositional structure and performance expression to the communication of emotion in music. *Psychology of Music, 42*(4), 503-524. https://doi.org/10.1177/0305735613482023

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Richins, M. L. (1997). Measuring emotions in the consumption experience. *Journal of Consumer Research, 24*(2), 127-146. https://doi.org/10.1086/209499

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178. https://doi.org/10.1037/h0077714

Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review, 3*(4), 444-454. https://doi.org/10.1177/1754073911410745

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1*(4), 331-346. https://doi.org/10.1007/BF00992539

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception: An Interdisciplinary Journal, 21*(4), 561-585. https://doi.org/10.1525/mp.2004.21.4.561

Shaver, P., Schwartz, J., Kirson, D., & O'Conner, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52*(6), 1061-1086. https://doi.org/10.1037//0022-3514.52.6.1061

Stoessel, J., Spreadborough, K., & Antón-Méndez, I. (2021). The metaphor of sweetness in medieval and modern musiclListening. *Music Perception: An Interdisciplinary Journal, 39*(1), 63-82. https://doi.org/10.1525/mp.2021.39.1.63

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition and Emotion, 22*(4), 720-752. https://doi.org/10.1080/02699930701503567

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

APPENDIX A: LIST OF MUSICAL EXAMPLES

Example label	Implied quadrant[1]	Quadrant rating in present experiment	Source
E1	Q1 (*playful*)	Q1/Q2	Heinichen, 1728: 362–364
E2	Q2 (*furious*)	Q1	Heinichen, 1728: 333–335
E4	Q4 (*love*)	Q1	Heinichen, 1728: 348–349
G03	Q1	Q1	Vieillard *et al.*, 2008: 746
P07	Q2	Q2	Vieillard *et al.*, 2008: 750
A07	Q4	Q3/Q4	Vieillard *et al.*, 2008: 745

E1 (*playful*)

Description: "…we could represent the result or the consequences of the search and believe that Aminta had found his love; then in this case the imagination takes the opportunity to portray the **playful** looks of <u>love</u>:" (Heinichen, 1728:362)[2]

E2 (*furious*)

Description: "Only now the composer can derive from Metilde's intentions that this in itself dry aria can be represented in the most **furious** of affections, which should fire invention-rich composers to

---

[1] For examples from Vieillard *et al.* (2008), implied quadrant refers to the mean ratings given by participants for each example in the experiment described in that paper. For examples from Heinichen (1728), implied quadrant refers to the quadrant attached to the keywords extracted from each example's accompanying verbal description, as described in the Methods section above. In the latter case, extracted keywords are given in parentheses.

[2] Underlined words represent candidate emotion terms. Words additionally in boldface are the emotion terms selected for each excerpt.

transform their formerly suspended thoughts into beautiful musical ideas. But should the natural fantasy require still more help, one can proceed to special expressions of the recitative such as: *alti dissegni, e precipizii immensi*, and these could give something like the following expression (or ten other inventions of this type):" (Heinichen, 1728:332)

E4 (*love*)

Description: "Should one wish to try special expressions, the *words faville, pupille, l'ardore, lo squardo* give our imagination much opportunity for pleasant and almost playful inventions. For example, one could represent the burning fire of **love** in the following invention:" (Heinichen, 1728:348)

G03

P07

A07

Note: Changes in time signature in the final bar of each example are artefacts of shortening the examples for use in the experiment, as described in the Methods.

APPENDIX B: PRECISE CUE LEVELS AND RATING DATA FOR ALL EXAMPLES

Example label	REP	Mean pitch height	Mean melodic interval size	Mode	Mean valence rating	Mean arousal rating	Median valence rating	Median arousal rating
E1	5.38	8.66	2.22	Minor	-0.13 (SD = 2.06)	1.5 (SD = 2.06)	0	1.5
E2	7.32	3.50	3.84	Major	3 (SD = 1.84)	3.43 (SD = 1.81)	4	4
E4	6.17	9.54	3.65	Major	3.63 (SD = 1.19)	3.13 (SD = 1.28)	4	3
G03	4.4	4.07	4.22	Major	3.97 (SD = 1.10)	3.1 (SD = 1.71)	4	3.5
P07	1.67	1.2	1.89	Minor	-1.2 (SD = 1.94)	1.23 (SD = 1.72)	-1	1.5
A07	2.3	-1.30	4.57	Major	0.13 (SD = 2.01)	-2.3 (SD = 1.27)	0	-2