Analysis of Emotional Dimensions in Music Using Time Series Techniques

Emery Schubert

Few who have experienced the thrill of emotional interactions with music would deny that such experiences are important and enriching. These intoxicating episodes can be superior to any heightening or soothing drug because they occur without unpleasant side-effects. The fact that music can express such emotions—move people to tears of lament and to teeth-clenching ecstasy—is, for many, a wondrous mystery. How is it that by manipulating combinations of innocuous acoustic parameters composers and performers can produce such potent effects? The question has occupied scholars and practitioners for over twenty centuries. One of the many problems that face researchers is the lack of objective data available for comparing music and emotion. Further, experimental research in this area has tended to treat music as a stimulus that can be dealt with by a single point response. While experimental methodology requires an element of reductionism, in this article I argue that it is possible to obtain an additional dimension to the understanding of emotion in music by tracking emotional responses to music continuously. My work is based on the fundamental assumption that there are underlying rules that govern the relationship between music and emotion within a given culture.[1]

The strength of the experimental approach has been summarised by Pike, who indicated the benefit of collecting responses from several people instead of the introspections of the

[1] My approach implies both a reductionist and behaviourist approach. While the approach of the reductionist is present, I rebuff the claim of a behaviourist approach. Behaviourists study input and output and their correlation and then contend that nothing else is worthy of study. See B.F. Skinner, *About Behaviorism* (Oxford: Alfred A. Knopf, 1974). Those who believe that this correlation is interesting (including this author) must still study input (music), output (emotion) and their correlations. In other words, I support the notion that there is a necessary cognitive mediation between music and emotion. As John Sloboda put it: 'Seen with the cold eye of physics a musical event is just a collection of sounds with various pitches, durations, and other measurable qualities. Somehow the human mind endows these sounds with significance. They become symbols for something other than pure sound, something which enables us to laugh or cry, like or dislike, be moved or be indifferent.' J.A. Sloboda, *The Musical Mind: The Cognitive Psychology of Music* (London: OUP, 1985) 1.

lone expert.[2] However, the empirical approach, stemming from the turn of the century, was often hindered by methodological flaws and technological deficiencies. Such studies provided fuel for the formalists and expressionists who believed in the strongly subjective nature of musical experience.[3] For example, Heinlein found that the widely accepted relationship between happy–major and sad–minor was without foundation, but some fifty years later the same data were reanalysed by Crowder and found to be misinterpreted.[4] In the meantime, several empirical investigations upheld the thesis of a relationship between emotions and musical features, even though this view received no serious attention from introspectionists.[5] In his influential book, *Emotion and Meaning in Music*, Meyer referred to the Heinlein study,[6] but not to the work of Hevner, Gundlach, Sherman, Rigg and several others who, in the 1920s and '30s, had found more positive relationships between *specific* emotions and musical features.[7]

The methodological problems that faced empirical researchers were numerous. 'Real' pieces of music usually have a large number of interacting musical features making the investigation of possible causal connections between musical features and emotional response more difficult to unravel. Further, some researchers required listeners to select a word from a list to describe the emotion in a piece of music, even though the piece typically expressed more than one emotion. This methodology made it impossible to distinguish whether the listener was making an overall reflection about the piece or responding to a salient part of the music.

One solution was to select isolated samples of sound so that a single musical structure or feature could be manipulated and tested for emotional expressivity. Such research is often criticised because it is too far removed from naturalistic music. While this may be true, it was one of the few practical ways of investigating the problem, and it could supply convergent evidence, or otherwise, with investigations that used 'real' music.

Until quite recently empirical researchers had to contend with static, asynchronous measurements of the dynamic process of music and emotional response. It is easy to see why this type of approach can be criticised: how could measured responses made at the end of a piece possibly cover the subtle and abrupt changes which music expresses during a performance? However, advances in computer technology now facilitate improvements in methodology that were not available to these early pioneers. Modern technology offers new opportunities

---

[2] A. Pike, 'A Phenomenological Analysis of Emotional Experience in Music,' *Journal of Research in Music Education* 20 (1972): 262–68.

[3] L.B. Meyer, *Emotion and Meaning in Music* (Chicago: U of Chicago Press, 1956); B. Reimer, *A Philosophy of Music Education* (Englewood Cliffs, NJ: Prentice Hall, 1989).

[4] C.P. Heinlein, 'The Affective Characters of the Major and Minor Modes in Music,' *Journal of Comparative Psychology* 8 (1928): 101–42 and R.G. Crowder, 'Perception of the Major/Minor Distinction: I. Historical and Theoretical Foundations,' *Psychomusicology* 4.1–2 (1984): 3–12.

[5] Meyer, *Emotion and Meaning in Music*; S.K. Langer, *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite, and Art* (Cambridge, MA: Harvard UP, 1957).

[6] Heinlein, 'Affective Characters.'

[7] K. Hevner, 'The Affective Character of the Major and Minor Modes in Music,' *American Journal of Psychology* 47 (1935): 103–18; K. Hevner, 'Experimental Studies of the Elements of Expression in Music,' *American Journal of Psychology* 48 (1936): 246–68; R.H. Gundlach, 'Factors Determining the Characterization of Musical Phrases,' *American Journal of Psychology* 47 (1935): 624–43; M. Sherman, 'Emotional Character of the Singing Voice,' *Journal of Experimental Psychology* 11 (1928): 495–97; M. Rigg, 'An Experiment to Determine how Accurately College Students can Interpret the Intended Meanings of Musical Compositions,' *Journal of Experimental Psychology* 21 (1937): 223–29.

to develop methods of obtaining responses often enough during a performance to capture the dynamic flow of expressions in the music.

Measurement of continuous response to music is a growing area of research, but such research has not been supported with quantitative analyses. Many previous music-emotion studies fail to use analytic techniques for relating continuous responses with the components of the continuous stimulus. Sophisticated time series analytic techniques have been available in the fields of statistics, geology and econometrics for more than thirty years, yet these procedures are only starting to spread to the study of emotion in music.

I have described my position with regard to the relationship between emotion and music: I believe that there is a relationship between emotion and music, and further, that there is an underlying, quantifiable relationship between musical features and emotional response. Before describing how this information can be modelled using time series analysis it is necessary to define the structure of emotion so that we can establish a means for quantifying it.

The Structure of Emotion

There are two broad systems of classifying emotions: categories and dimensions. Categorical classification of emotion assumes that emotions carrying different meanings, such as happy and sad, are distinct and independent entities. Checklists imply a categorical classification of emotions. For example, Tomkins's research on facial expressions suggested that emotions could be grouped into one of eight categories:[8]

1. interest/excitement
2. enjoyment/joy
3. surprise/startle
4. distress/anguish
5. fear/terror
6. shame/humiliation
7. contempt/disgust
8. anger/rage.

In contrast, the dimensional classification of emotion holds that all emotions are in some way related within an n-dimensional *semantic space* (or, more correctly, *emotion space)*. For example, the dimensional structure suggests that happy and sad are opposite emotions along the valence dimension of emotion. To distinguish distinct emotions having similar valence, such as *sad* and *angry*, a second dimension, *arousal* (or *activity*), may be added: sad (typically) has low arousal and angry has high arousal.

Dimensional classification can help to visualise the interrelationships between emotions and to provide a structured framework for emotion research. Several prominent psychologists support the use of the dimensional approach for the study of emotions.[9] However, opinions on

---

[8] S.S. Tomkins, *Affect, Imagery, Consciousness* (New York: Springer, 1962); S.S. Tomkins and C.E. Izard, eds, *Affect, Cognition, and Personality: Empirical Studies* (New York: Springer, 1965).

[9] H. Schlosberg, 'Three Dimensions of Emotion,' *Psychological Review* 61 (1954): 81–88; R. Plutchik, *The Emotions: Facts, Theories and a New Model* (New York: Random House, 1962); J.A. Russell, 'A Circumplex Model of Affect,' *Journal of Social Psychology* 39 (1980): 1161–78; M.A. Zevon and A. Tellegen, 'The Structure of Mood Change: An Idiographic/Nomothetic Analysis,' *Journal of Personality & Social Psychology* 43.1 (1982): 111–22; J.A. Russell, A. Weiss and G.A. Mendelsohn, 'Affect Grid: A Single-item Scale of Pleasure and Arousal,' *Journal of Personality and Social Psychology* 57.3 (1989): 493–502; P.M. Niedenthal and M.B. Setterlund, 'Emotion Congruence in Perception,' *Personality & Social Psychology Bulletin* 20.4 (1994): 401–11; J.A. Russell, 'How Shall an Emotion be Called?' *Circumplex Models of Personality and Emotions*, ed. R. Plutchik and H.R. Conte (Washington: American Psychological Association, 1997): 205–20.

the number of dimensions and the definitions of these dimensions do not always converge.[10] Whissell and associates asserted that there has been considerable agreement on the validity of the dimensional paradigm of emotion consisting of two dimensions, namely valence and arousal.[11] Whissell and her colleagues also pointed out that there has been considerable disagreement on 'the role which any other dimensions may play (attention, competence, locus of causation, potency, dominance, have all been suggested as additional dimensions)' and that the first two dimensions of valence and arousal can explain up to eighty percent of response variance.[12] Potency, a dimension believed by some researchers to be the best candidate for a third dimension, appeared to explain less than five percent of the variance according to Whissell.[13] Sweeney and Whissell suggested that the valence and arousal dimensions were 'theoretically interpretable' as well as being distinct.[14]

Over several publications, Whissell and associates reported the development of a 'dictionary of affect.' In one study by Whissell et al., participants rated lists of words on two seven-point, bipolar scales, one for 'evaluation (pleasantness)' and another for 'activation (arousal).'[15] This methodology enabled faster responses compared with other response measures available (two hundred words in less than thirty minutes, or less than nine seconds per word) within a plausible paradigm of emotions (using only the two most salient dimensions).

It is conceivable that Whissell and her associates may have thought to have responses made directly in an emotion space. That is, the arousal and valence bipolar scales could be combined at right angles enabling simultaneous response to the two dimensions. Such a format may have aided in speeding up responses and provided a convenient visualisation of the procedure for the participant. But with large lists of words on a paper and pencil test, an emotion space may not have been practical. Apart from Russell's affect grid,[16] no studies have been cited before my work in which an emotion space was used as a direct response measure.[17] However, such a development was considered pivotal to the data gathering and analytic approach I am presenting here.

Russell provided potent evidence about the merit of the dimensional system of classification and in particular the circumplex realisation upon a two-dimensional emotion space consisting

---

[10] R.J. Larsen and E. Diener, 'Promises and Problems with the Circumplex Model of Emotion,' *Emotion Review of Personality and Social Psychology* 13, ed. M.S. Clark (Thousand Oaks, CA: Sage Publications, 1992): 25-59; J.S. Roberts and D.H. Wedell, 'Context Effects on Similarity Judgments of Multidimensional Stimuli: Inferring the Structure of the Emotion Space,' *Journal of Experimental Social Psychology* 30.1 (1994): 1–38; U. Schimmack and A. Grob, 'Dimensional Models of Core Affect: A Quantitative Comparison by Means of Structural Equation Modeling,' *European Journal of Personality* 14.4 (2000): 325–45.

[11] C. Whissell and H. Berezowski, 'A Dictionary of Affect in Language: V. What is an Emotion?' *Perceptual & Motor Skills* 63.3 (1986): 1156–58; C.M. Whissell, M. Fournier, R. Pelland, D. Weir and K. Makarec, 'A Dictionary of Affect in Language: IV. Reliability, Validity, and Applications,' *Perceptual & Motor Skills* 62.3 (1986): 875–88; R. Whissell and C. Whissell, 'The Emotional Importance of Key: Do Beatles Songs Written in Different Keys Convey Different Emotional Tones?' *Perceptual & Motor Skills* 91.3 (2000): 973–80.

[12] Whissell et al., 'Dictionary of Affect in Language' 876.

[13] Whissell et al., 'Dictionary of Affect in Language' 876.

[14] K. Sweeney and C. Whissell, 'A Dictionary of Affect in Language: I. Establishment and Preliminary Validation,' *Perceptual and Motor Skills* 59 (1984): 695–98.

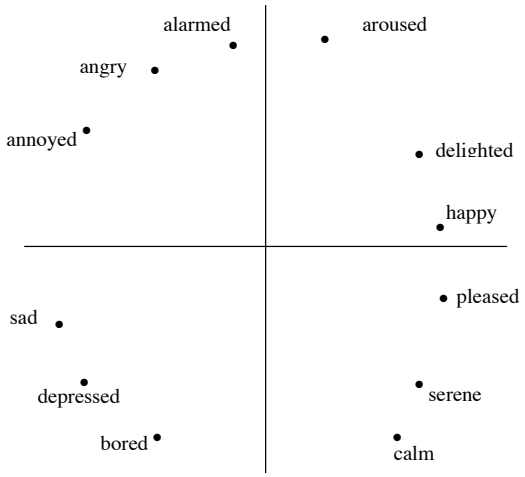
[15] Whissell et al., 'Dictionary of Affect in Language' 877.

[16] Russell, Weiss and Mendelsohn, 'Affect Grid.'

[17] E. Schubert, 'Measuring Temporal Emotional Response to Music using the Two-dimensional Emotion Space,' *Proceedings of the 4th International Conference for Music Perception and Cognition* (1996): 263–68.

Figure 1. Circumplex model of emotion-related categories

The circumplex model of emotion demonstates how emotions can be plotted on a two-dimensional space. Examination of the plot demonstrates the interrelationship of emotions. Emotions plotted higher on the space have higher arousal (such as alarmed and aroused). Emotions lower on the emotion space have low arousal (such as bored, sleepy and calm). Emotions to the right are positive (pleased, happy) and emotions to the left are negative (sad, annoyed, depressed). Emotions plotted closer together have more similar meanings than those further apart (for example, delighted and happy are close, but delighted and depressed are distant). The emotion space therefore provides a means of quantifying aspects of emotion, whether in terms of meaning expressed by words (as shown in the figure), or faces, or music. The sample words shown are based on J.A. Russell, 'Measures of Emotion,' *Emotion: Theory Research and Experience 4* ed. R. Plutchik and H. Kellerman (New York: Academic Press, 1989) 86.



of valence and arousal.[18] In such a configuration emotions line up roughly along a circle centred on the cartesian plane (see Figure 1). According to the circumplex model, words geometrically close together, such as delighted and happy, indicate closeness in meaning. Other researchers have used the emotion space to represent non-verbal stimuli. Schlosberg asked participants to rate emotion expressed by faces on two scales: pleasantness and attention-rejection. Responses were then plotted on a two-dimensional emotion space.[19]

In recent studies dealing with emotional responses to music, the dimensional interpretation of data has pervaded the literature. Gundlach's analytic approach pioneered the dimensional interpretation of emotional data collected in response to music,[20] however the techniques of multidimensional clustering became firmly rooted in music-emotion research from the early 1970s.[21] Many of the studies agreed or implied that the first two dimensions were those associated with valence and arousal, thus supporting the use of the valence and arousal

[18] J.A. Russell, 'Measures of Emotion,' *Emotion: Theory Research and Experience* 4, ed. R. Plutchik and H. Kellerman (New York: Academic Press, 1989) 81–111.

[19] H. Schlosberg, 'The Description of Facial Expressions in Terms of Two Dimensions,' *Journal of Experimental Psychology* 44 (1952): 229–37.

[20] Gundlach, 'Factors Determining the Characterisation of Musical Phrases.'

[21] See L. Wedin, 'Evaluation of a Three-dimensional Model of Emotional Expression in Music,' *Reports From the Psychological Laboratories, University of Stockholm* 13 (1972): 241–57; L. Wedin, 'A Multidimensional Study of Perceptual-emotional Qualities in Music,' *Scandinavian Journal of Psychology* 13.4 (1972): 241–57.

dimensions of emotion as measures of response to music in addition to other modes of stimuli. This same two-dimensional representation of emotion is common in current music perception research.[22]

Also, a strong intuitive implication is that the construct of emotion by which stimuli such as words and pictures are judged is similar, if not the same, as the construct used to judge emotion expressed by music. There appears no obvious reason why the emotional conception of sadness expressed by a face would be different to the emotional conception of sadness expressed by a piece of music.[23] To put it in another way, emotional concepts appear to be stable across various modes of perception.

Studies of multidimensional response to music have required the participant to make responses about an entire selection with several or even dozens of response items. In contrast, more recent research has focused on continuous response using a single dimension to measure response.[24] The continuous response digital interface (CRDI) is a specially designed instrument commonly used to collect such data, with numerous published studies resulting.[25] Many of these unidimensional studies have deliberately left the meaning of the dimension measured subjective and vague, however given their musical correlates, I have speculated that a construct related to the arousal dimension was being measured.[26] A natural progression from the work of those measuring continuous response to music is to increment the number of dimensions available for response. By allowing the participant to respond along two dimensions instead of one, the effects of *cognitive load* will be increased minimally, and the two important dimensions of emotion may be employed.[27]

In 1996 I presented data from a software program that collected emotions continuously in response to music via a two-dimensional emotion space. At around the same time, Clifford Madsen was doing the same with the development of the two-dimensional CRDI.[28] The data

---

[22] See, for example, C.C.V. Witvliet, The Impact of Music-prompted Emotional Valence and Arousal on Self-report, Autonomic, Facial EMG, and Startle Responses across Experimental Contexts, PhD thesis, Purdue University, 1992; Whissell and Whissell, 'Emotional Importance of Key'; N. Dibben, 'The Role of Peripheral Feedback in Emotional Experience with Music,' *Music Perception* 22.1 (2004): 79–115; P. Gomez and B. Danuser, 'Affective and Physiological Responses to Environmental Noises and Music,' *International Journal of Psychophysiology* 53.2 (2004): 91–103.

[23] For further discussion on this matter see P. Kivy, *Introduction to a Philosophy of Music* (Oxford: Clarendon, 2002).

[24] A. Goldstein, 'Thrills in Response to Music and other Stimuli,' *Physiological Psychology* 8 (1980): 126–29; F.V. Nielsen, 'Musical Tension and Related Concepts,' *The Semiotic Web '86: An International Year-book*, ed. T.A. Sebeok and J. Umiker-Sebeok (Berlin: Mouton de Gruyter, 1987); C.K. Madsen and W.E. Frederickson, 'The Experience of Musical Tension: A Replication of Nielsen's Research using the Continuous Response Digital Interface,' *Journal of Music Therapy* 30.1 (1993): 46–63; M. Waterman, 'Emotional Responses to Music: Implicit and Explicit Effects in Listeners and Performers,' *Psychology of Music* 24.1 (1996): 53–67; J.A. Sloboda and A.C. Lehmann, 'Tracking Performance Correlates of Changes in Perceived Intensity of Emotion During Different Interpretations of a Chopin Piano Prelude,' *Music Perception* 19.1 (2001): 87–120.

[25] For an overview, see <http://musictherapy.fsu.edu/crdi/referencetable.html>.

[26] E. Schubert, 'Continuous Measurement of Self-report Emotional Response to Music,' *Music and Emotion: Theory and Research*, ed. P. N. Juslin and J.A. Sloboda (New York: OUP, 2001): 393–414.

[27] D.A. Norman and D.G. Bobrow, 'On Data-limited and Resource-limited Processes,' *Cognitive Psychology* 7.1 (1975): 44–64; J. Sweller and P. Chandler, 'Evidence for Cognitive Load Theory,' *Cognition & Instruction* 8.4 (1991): 351–62; J.J. van Merrienboer and J. Sweller, 'Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions,' *Educational Psychology Review* 17.2 (2005): 147–77

[28] C.K. Madsen, 'Emotion versus Tension in Haydn's "Symphony no. 104" as Measured by the Two-dimensional Continuous Response Digital Interface,' *Journal of Research in Music Education* 46.4 (1998): 546–54.

from these devices show moment-to-moment shifts of emotion as the music unfolds. Such data were subsequently shown to produce a reliable measure of the typical emotional response which music expressed over time as according to a typical listener. I will now describe how these data can be used to help understand the way in which musical features can be combined to produce aspects of emotional responses. My goal is to head toward a formal emotional analysis of music.

Defining Musical Features

Musical features are the separable elements of music or the perceptually distinguishable combinations of elements which, when combined, form a musical object. Based on the work of Seashore, these can be defined at various levels: a low, psycho-acoustic level, or a high, formal-structural level.[29] Low-level musical elements consist of pitch, loudness, timbre and duration. These elements are the universal constituents of musical sound, at least from a western perspective. This makes them necessary elements in the construction of music.[30]

High-level elements comprise such features as harmony, voicing, phrasing, texture, form and style. These higher-level elements, although specifically related to music, are often culturally specific and contain an element of subjectivity. In fact, these higher-level features are in some cases peculiar to Western music. For example, classical western music systems of harmony are quite different from Eastern systems.[31] Nevertheless, within a cultural context these musical features are meaningful, perceptually valid components of music.

Somewhere in between these two extremes lie rhythm, contour, envelope and articulation. The levels of musical features that I have proposed here are not always clear, nor are they always necessary (see note 30). It is, however, important to note that all of the higher-level musical elements can be expressed in terms of the lowest, sound-defining elements. For the sake of simplicity, I will begin by restricting analysis to a selection of more or less low-level musical features: pitch, loudness, tempo, timbral brightness and texture.

Relating Musical Features and Emotions: A Tutorial in Analysing Emotion in Music using Autoregression Modelling

We are presently interested in how variations in musical features can be used to predict emotional response. A statistical model that can be used to express such a relationship is called regression modelling. Regression models provide a mathematical representation of how one variable can predict another. We apply regression-type formulas in their simplest sense when we are shopping, and can calculate how much two toilet rolls cost when the display on the aisle indicates the cost of one toilet roll. The cost of the toilet roll predicts how much each roll will cost (obviously). So two toilet rolls will cost twice as much as one toilet roll. Because we can simply multiply the price of one roll by the number of rolls (rather than doing something more complex), we can call this a linear model. There are no statistics involved in the toilet

[29] C.E. Seashore, *Psychology of Music* (New York: McGraw Hill, 1938/1967).

[30] Actually, frequency and intensity are the two necessary and sufficient acoustical parameters of all music; but I use the weaker definition because it is conceptually simpler.

[31] See, for example, M.A. Castellano, J.J. Bharucha and C.L. Krumhansl, 'Tonal Hierarchies in the Music of North India,' *Journal of Experimental Psychology: General* 113 (1984): 394–412.

roll example because we know the exact price of each unit. However, suppose we wanted to predict how arousing a piece of music is by using just loudness. It might well be that doubling the loudness will double the emotional arousal produced. But the story becomes a bit more complicated. First we need examples of data with a given arousal value and a given loudness level. To ascertain the arousal level for a given piece of music we might ask a number of participants to rate arousal for the given piece and take their averaged response. There are several ways of measuring loudness, and this can be done relatively easily. So in the case of continuos data we have changing arousal values and changes in loudness that occur at the same time. From these data we can calculate a relationship between loudness and arousal. Indeed, this can and has been done, and loudness alone has been shown to be a reasonably good predictor of emotional arousal expressed by music.[32] However, it should also be clear that this is a major simplification. Firstly, let us recall the assumption that there is an underlying relationship between loudness and arousal. In measuring arousal and loudness mistakes are made: participants can vary in their responses (both within their own response, and with respect to others), and the measurement of loudness is itself not going to be a precise representation of every person's perception of loudness. These variations produce an error in the prediction, and with this 'error term' we have the basis of linear regression.

While avoiding the details of the derivation of regression equations, it is necessary to understand one further complication in those regression models that attempt to predict a time-dependent response. When a piece of music unfolds, our responses are not solely dependent on the musical features being sounded at the point in time where the emotional response is being made. That is, if we could momentarily freeze the continuous responding of our typical listener, the situation is more complex than simply looking at the musical features at that same point in time. First, we will ignore cultural variables, individual differences and the mood of the listener, not because they are unimportant, but because we need to simplify what is already a fairly complex problem. The complexities we shall examine are those of the short-term memory of the listener. This means that the listener is not only processing music as it is heard, but also that the listener stores information about what has unfolded in the music. When this time dependency is incorporated into our regression equation we are examining one kind of time-series model, the 'autogression' model. The mathematical details of autoregression models can be found elsewhere,[33] however it may be understood conceptually as adding memory to the model: as the music unfolds, the emotional response is evaluated by the listener as being some combination of emotional responses and musical features from an earlier time in the piece.

In statistics this time dependency is referred to as *serial correlation* and is a critical factor in most time-dependent phenomena. For example, when we walk down the street, our position at any given time can be well predicted by where we were at the time we took our previous step. It can also be predicted by the position we were at the time of the step before that, but possibly not as well. Where we were the day before the walk is going to be a much poorer

---

[32] E. Schubert and W. Dunsmuir, 'Regression Modelling Continuous Data in Music Psychology,' *Music, Mind, and Science*, ed. S.W. Yi (Seoul: Seoul National U, 1999) 298–352.

[33] G.E.P. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control* (San Francisco: Holden-Day, 1976); J.M. Gottman, *Time-Series Analysis: A Comprehensive Introduction for Social Scientists* (Cambridge: CUP, 1981); C.W. Ostrom, *Time Series Analysis Regression Techniques* (Newbury Park, CA: Sage, 1990); Schubert and Dunsmuir, 'Regression Modelling.'

predictor of the current position. That is, as we go back in time we can make predictions about where we are now, but the prediction gets worse as we go further back. The analogy in the cognitive domain is memory. Our short-term memory has a limited capacity to store information, but still enough to allow us to remember recent information about the music to which we are attending, whether consciously or otherwise. So when we are making emotional responses or interpretations continuously, the response will be a function what has gone on before in the musical features.[34]

Case Study 1: Grieg 'Peer Gynt'

An example of a regression model that predicts arousal was presented by me using, among other pieces, 'Morning' from the *Peer Gynt Suite*, op. 46, no. 1, by Edvard Grieg.[35] I will unpack this example to demonstrate how an emotional analysis could be conducted for one of the two emotional dimensions discussed above: arousal. The model used not only loudness, but tempo, timbral brightness, melodic pitch and texture to predict emotional response. Because people listen and then respond, there is likely to be some small delay between the musical features and the resultant emotional response. Therefore, the musical feature occurring for several seconds before the emotional response was included in the regression model. The arousal response to the Grieg could be expressed conceptually as:

Arousal = loudness (3) + tempo (3) – brightness (4).

The number in parentheses refers to the number of seconds of lag in arousal response before that feature could best predict the response. The musical features are listed in order of predictive strength of arousal. So, the strongest predictor of emotional arousal response to the Grieg is made by changes in loudness, and in particular when the change in loudness occurs three seconds before the response. Loudness is by far the strongest predictor of arousal response. The next strongest predictor of arousal response is tempo. So changes in tempo affect arousal response also with a delay of about three seconds. Timbral brightness has a weaker effect on arousal, and the effect occurs about four seconds after the change in brightness occurs. The negative sign before brightness indicates that arousal changes in the opposite direction to changes in brightness. So, as the tone of the orchestra becomes brighter, the arousal decreases. As the tone of the orchestra becomes less bright, arousal increases, but the effect is marginal.

As shown in the Figure 2a, arousal starts off quite low in the piece, until about the fiftieth second has elapsed, at which time a large crescendo occurs with the *tutti* orchestra (Figure 2b). Visual inspection of the arousal and loudness curves shows a fair amount of similarity in movement, and in each case the arousal response occurs a little bit after the change in loudness. For example, by the fifty-first second, loudness has reach one of the highest points of the piece, whereas arousal is just starting to rise.

---

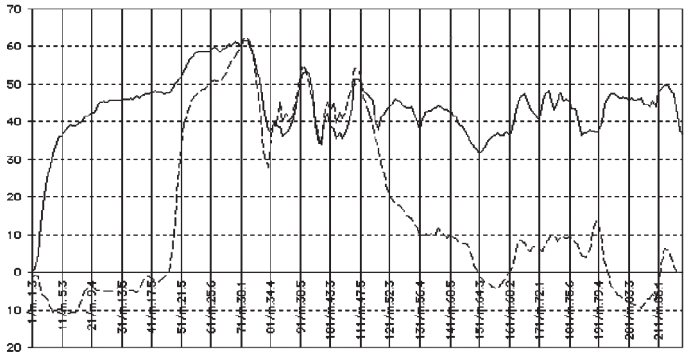
[34] Memory and psychological theories of memory are more complex than is implied by this simple explanation. For more information see J.R. Anderson, 'A Spreading Activation Theory of Memory,' *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*, ed. A.M. Collins and E.E. Smith (San Mateo, CA: Morgan Kaufmann, 1988) 137–54; J.R. Anderson and L.J. Schooler, 'The Adaptive Nature of Memory,' *The Oxford Handbook of Memory*, ed. E. Tulving and F.I.M. Craik (London: OUP, 2000): 557–70.

[35] Using the performance by CSSR State Philharmonic Orchestra, conducted by Stephen Gunzenhauser. See E. Schubert, 'Modeling Perceived Emotion with Continuous Musical Features,' *Music Perception* 21.4 (2004): 561–85.

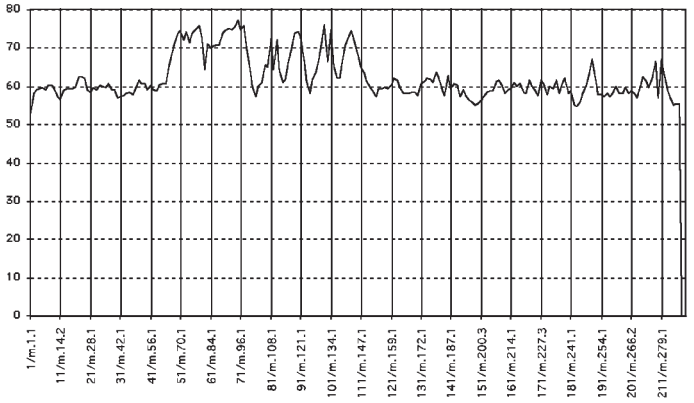
Figure 2. Time series plots for 'Morning'

The x-axis displays the time elapsed and the corresponding bar number in the form '51/m.21.5' where 51 is the time elapsed in seconds and 21.5 is the fifth beat of the 21st bar (in 6/8 metre). Duration 3 minutes and 38 seconds (218 seconds) for the performance by CSSR State Philharmonic Orchestra, conducted by Stephen Gunzenhauser, *Discover Classical Music*, CD 2, Naxos NHN 8.550009 (1993).

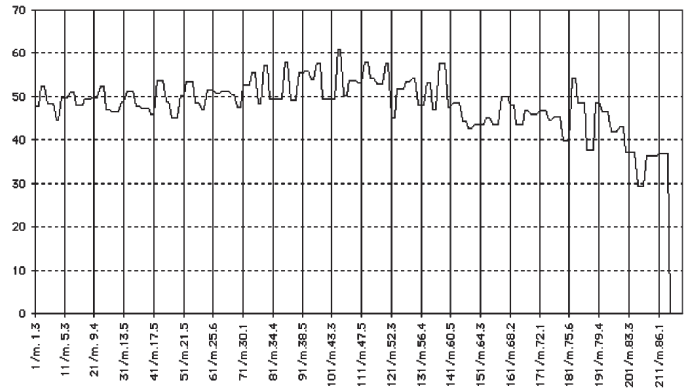
a) typical valence % (filled line) and arousal % (dashed line) responses



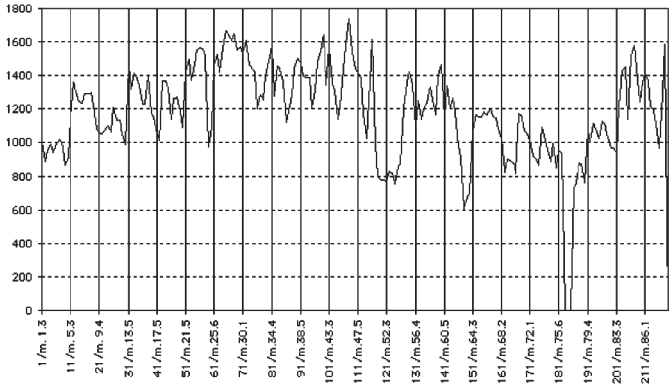
b) loudness



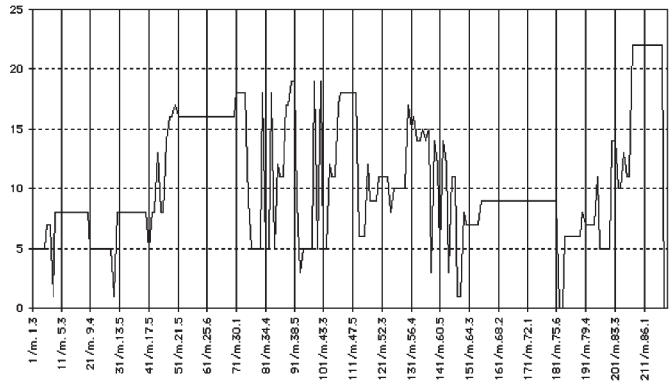
c) melodic pitch (note number, where 60 is middle C, 72 is the C an octave above)



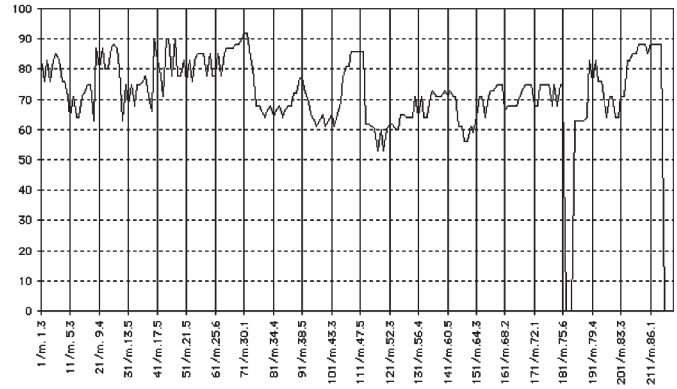
d) centroid in Hz (or 'timbral brightness')



e) texture (number of instruments playing)



f) tempo in bpm



The effect of tempo is more subtle. The tempo commences at about fifty beats per minute (bpm) (see Figure 2f). While fluctuations in the tempo curve are prominent, indicating phrase shapes, there is an overall arch pattern, with an increase to around 55 bpm between bars 30 and

52, peaking at over 60 bpm in bar 45, and gradually declining, reaching 30 bpm in bar 85, but with a slight peak at bar 77 (about 54 bpm). It can be seen that the arousal rises dramatically from around the fiftieth second, and then declines most dramatically from the 111th second. This region corresponds to the increase in tempo. It is evident here that when loudness is altered, tempo too changes, even if not noticeably or consciously. And so it becomes clear why these two features are the strongest predictors of arousal response.

Much weaker is the effect of timbral brightness. Timbral brightness is measured by calculating the centre frequency (or, more formally, the frequency spectrum centroid, or 'centroid' as displayed in Figure 1d) of the sound spectrum at each point in time. Higher values indicate a brighter timbre.[36] The units of this centroid are in vibrations per second, or Hertz (Hz), but are most easily interpreted as a perceptual quality of the overall brightness in tone colour. It is difficult to see how a relationship can occur between arousal and brightness from the chart. Indeed, the relationship is quite weak. An investigation of the small dips that occur in brightness from bar 23 through to 52 provide some clues as to how the relationship comes about. The first dip, in the second half of bar 24, is due to the descending C♯, A, G♯ quavers in the contrabass and bassoon sections; the dip over bars 32 to 33 corresponds to a cello solo with bassoon and violin accompaniment (introducing the second theme), and this section is repeated using the same instrumentation with modulation over bars 40 and 41. The first major dip occurs at bar 50 which is due to the prominent, mellow third horn playing the first theme with *pianissimo* semiquaver arpeggios in the upper woodwinds, *pizzicato* upper strings, and sustained cello and contrabass accompaniment. Each of these dips occurs in a region of high arousal. That is, the brightness seems to move at times in the opposite direction to the arousal; subsequently, the coefficient for brightness term in the equation is negative.

The texture, which is simply coded as the number of instruments playing at any time, does seem to resemble the arousal pattern, particularly from the beginning where the orchestration is quite reduced (interchanging flute and oboe solos) until the crescendo with full orchestra at bar 21. Again, we can see a shared relationship between loudness and texture (as we saw with loudness and tempo). However, there are occasions in the piece when texture is quite high, but arousal is quite low, such as the end of the piece (from bar 85), where the orchestra is playing *tutti*, but very softly. This kind of mismatch between texture and loudness provides some strong evidence of the power of loudness (beyond texture) to manipulate perceived arousal, and explains why texture did not play an important role in arousal for this piece. Melodic pitch is another variable that did not make a statistically significant contribution to arousal.

One parameter that has not been identified in the analysis of this regression equation is the serial correlation, or memory effect described earlier. Put simply, the autocorrelation term models the amount of information that is retained from the previous step (as in the previous moment in the music). In the model under scrutiny, this information is updated once each second (although this 'sampling rate' is more or less arbitrary, but should be no slower than once per second). The coefficient for the parameter is 0.51, on a scale of 0 to 1. This value indicates that a fair amount of information (crudely, half the emotional information) is remembered and incorporated into the response at the next moment (second) in time. While this may not seem

[36] E. Schubert and J. Wolfe, 'Does Timbral Brightness Scale with Frequency and Spectral Centroid?' *Acta Acustica united with Acustica* 92.5 (2006): 820–25.

terribly meaningful because it is not referring to a musical parameter, it is a crucial part of the autoregression time series model. It enables the model to explain an aspect of time dependency, which is an essential detail of the listening experience.

Case Study 2: Dvořák 'Slavonic Dance'

Using a similar approach to that described for 'Morning,' a statistical model can be produced for the arousal response to Antonín Dvořák's 'Slavonic Dance,' op. 46, no. 1.[37] For this model, 73 percent of the arousal response could be explained as according the conceptual formula:

$$\text{Arousal} = \text{loudness (0 to 4)} + \text{brightness (0 to 2)}.^{38}$$

Here, two musical parameters are sufficient to explain this large amount of variation in arousal. That is, arousal can be predicted by the instant at which loudness changed, one second after loudness changed, two seconds after loudness changed, and so on, up to four seconds after loudness changed. Similarly the brightness (or centroid) change predicted arousal over a span of three seconds (instantaneous through to two seconds after the change in brightness).

As found in the Grieg example, loudness seems to be the dominating parameter in determining arousal response. For the present example, two questions come to mind. Why does loudness now influence arousal instantaneously and why does brightness affect arousal?

As can be seen in Figure 3, the arousal curve follows the shape of both the loudness and the centroid profiles. We must proceed cautiously when interpreting this graph because the structure of the piece is more strophic than the Grieg. With repeating sections (for example, as shown in Figure 3, the first, rousing *furiant* theme returns at the fortieth second, at the 128th second in a shortened form, and again as part of the coda at the 211th second), the arousal, loudness and centroid each peak at around these points. However, since some of this material is repeated in the same key and with the same orchestration, it may give an impression of a regularity that is merely an artefact of a repeated pattern. So, if a slight relationship is found between centroid and arousal in one section, the relationship is reinforced simply through the repetition. Inspection of the score at those points also demonstrates another problem with the model. In nearly all of the loud sections, the upper strings enter, along with the cymbals. In effect, the tone colour of the orchestra becomes brighter, as reflected in the centroid plot in Figure 3, at each occurrence of the *furiant* theme. It could be that the composer is increasing brightness as a means of increasing loudness. While the converse may also be true (that a brighter sound was required by the composer, producing a louder sound), the point is that one could be seen as being statistically redundant in producing arousal.

Consider a thought experiment. Supposing Dvořák really did want to manipulate the arousal expressed by the piece. Were he asked to achieve this manipulation through loudness, while keeping centroid more or less constant (for example, take the violins and flutes down an octave, and remove the cymbal crashes and piccolo in the *furiant* sections, or maybe even keep the orchestration identical throughout), would he have been able to do this? Certainly he could use other means to manipulate arousal, such as tempo but, as the Grieg example

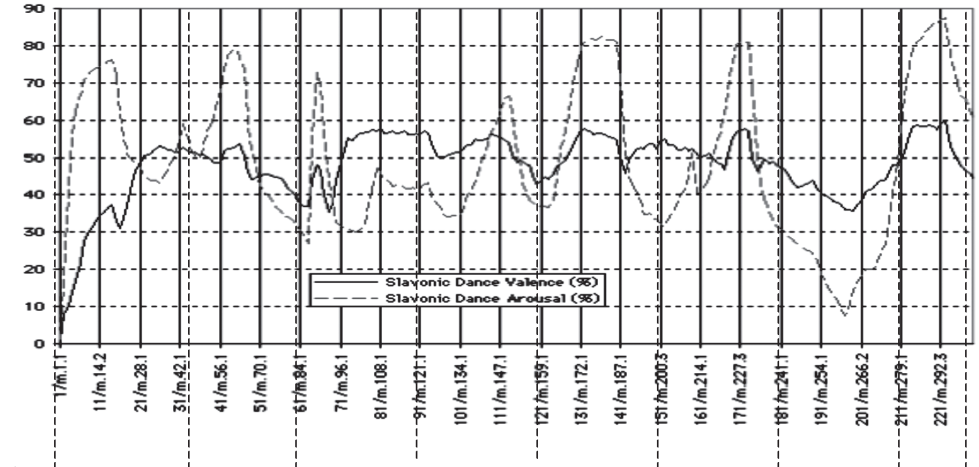
[37] Based on a performance by the Slovak Philharmonic Orchestra, conducted by Zdeněk Košler, *Discover Classical Music*, CD 2, Naxos NHN 8.550009 (1993).

[38] For more details, see Schubert, 'Modeling Perceived Emotion.'

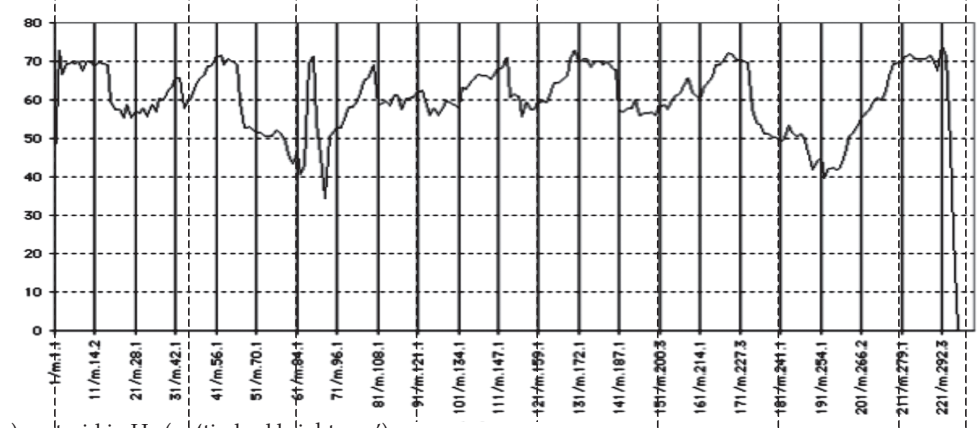
Figure 3. Time series plots for 'Slavonic Dance'

The x-axis displays the time elapsed and the corresponding bar number in the form '51/m.70.1' where 51 is the time elapsed in seconds and 70.1 is the first beat of the 70th bar. Duration 3 minutes and 45 seconds (225 seconds) for the performance by the Slovak Philharmonic Orchestra, conducted by Zdeněk Košler, *Discover Classical Music*, CD 2, Naxos NHN 8.550009 (1993).

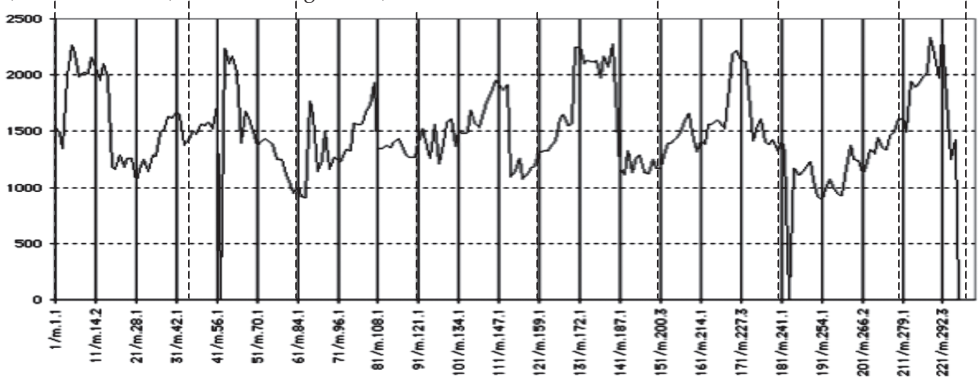
a) typical valence (filled line) and arousal (dashed line) responses



b) loudness



c) centroid in Hz (or 'timbral brightness')



demonstrated, loudness seems to be very important in the expression of arousal by the music. The brightness might not be providing any new information. In statistics this redundancy is referred to as multicollinearity. While this effect may have been slight, the repeating of the sections might have unfairly inflated the effect of brightness, as discussed.

I do not argue that brightness is therefore *musically* a redundant parameter. Composers often manipulate more than one feature in a united, dependent way. Having choices to achieve subtly different effects is one of the reasons why music is such an enriching art form. Understanding how these features each effect the listener's judgement is the purpose of this initial, if crude, attempt at analysing the emotion in the music.

Before concluding, I will discuss one more aspect of the Dvořák arousal model, namely the instantaneous loudness-arousal response. By this, I refer to the fact that the model predicts a change in arousal within one second of a change in loudness, unlike the Grieg that indicated a delay of around three seconds. Looking at the score, it is quite easy to see how this comes about. The score contains several *subito forte* effects, such as the sudden, abrupt loud chord with which the piece opens: notice in Figure 3 that loudness peaks at this very opening chord, and arousal rushes up to meet it very quickly. The sudden spasm of loudness at the 63rd second (surrounded by very quiet, lightly orchestrated sections) is another example. And again, the arousal rises almost as suddenly at around that time. So does valence, but valence does not produce anywhere near the number of regularities in response as a function of these variables as does arousal. A closer, statistical analysis was conducted on these suddenly changing fragments and it was concluded that it was at several of these sudden bursts of loudness that arousal response shifted to occurring sooner than was the case in other parts of the same piece, and in 'Morning'—which consisted of more gradual *crescendi*.[39] This unveils a further complication to the emotion model: the response time is not constant, but varies—what I would like to call a 'dynamic lag structure.' So, in some cases people notice a change in loudness and judge an arousal response some three seconds later, whereas in cases where the loudness changes are sudden, the corresponding arousal response occurs much faster, as though a startle or reflex action. Once again, it is this complexity that helps to make music such a rich experience. It may even be that dynamic lag structure is an aesthetic goal of the composer and performer, an attempt to alter the speed with which a listener responds to or experiences the music.

The autoregression coefficient for the Slavonic dance arousal model is 0.49.[40] Again, this means that approximately half of the previous value in arousal is having an influence on the arousal at the next point in time throughout the piece.[41] This effect propagates through the listening experience, as would one's ongoing memory of the unfolding piece: pieces closer to the present have a strong influence, but the effect becomes progressively weaker as we go further back in time.

Conclusion

My approach of modelling emotion as a continuous, statistical function of musical parameters in some ways harks back to the idea that music can be codified to have quite specific emotional

[39] Schubert and Dunsmuir, 'Regression Modelling.'

[40] Schubert, 'Modeling Perceived Emotion.'

[41] More correctly the 'arousal error term,' though the additional complexity of this concept is beyond the scope of this article. For more information, see Schubert and Dunsmuir, 'Regression Modelling.'

meanings: a revisitation of the so called doctrine of the affections, perhaps.[42] So within the tradition of trying to explain emotion in music, I argue that the statistical modelling of memory is a significant step forward in understanding aesthetic responses to music. While matters of culture, individual differences and the present state of the listener are factors which are sometimes ignored in music perception research but considered more by social scientists and cultural studies researchers, the essential contribution of memory in emotional response to music seems to have been ignored by most disciplines, particularly in terms of quantification.

The regression method has identified a subset of musical features that are able to predict emotional response, in the present case the arousal component of emotion perceived. The models explain responses of an idealised, typical listener, as represented by the sample who made their continuous responses to these pieces of music. More than sixty people gave their responses to the pieces. Most of them were musically experienced listeners and covered a wide range of ages. Of course any individual's response could be quite different from the typical listener's response. However, the models described here for the typical listener explain over sixty percent of the variation in response. That means that with just the five musical parameters of loudness, tempo, timbral brightness, texture and melodic pitch (the latter two making a negligible contribution to the prediction of arousal in both pieces investigated) we are able to fairly successfully predict the typical listener's arousal response. This figure would almost certainly be much lower for any particular individual. But the point is that the model could already make strong predictions without examining higher level musical features such as harmony.

Increasingly sophisticated time-series models will get better at representing human emotional response to music. This bottom up approach is compatible with traditional musicological views. The reason for a lacking nexus (if there is one) is due, more than anything else, to a 'hole in the middle.' With enough knowledge from empirical epistemologies about the quantifiable relationships between emotion and music, we might be able to test some of the more rationalist approaches that focus on top down matters. Consequently, personality, mood and perhaps even some cultural factors could be coded in a way meaningful enough to enable the modelling of a typical and perhaps even specific individual's emotional responses. This extreme level of reductionism might leave me labelled a perverse positivist, or perhaps with the accusation of trying to revive a doctrine of the affects through 'equations of the emotions.' I must therefore point out that music perception researchers interested in examining the relationship between musical features and emotional responses are aware that it is a complex problem, and that initially at least, they are seeking some simple, underlying regularities in response. Indeed, there can be no doubt that musical features can be used to predict emotional response to some degree. Therefore, the logical extension of this verifiable assumption is that the entire emotional experience could be modelled as a function of intra and extra musical factors. How this can be done is a matter for speculation, and I would be more content with a logical positivist label in aiming for such an ideal. However, in this article I have attempted to demonstrate conceptually how simple techniques available in several fields of research can be called upon to provide new points of departure in investigations of this fascinating problem of emotion in music.

---

[42] G.J. Buelow, 'Rhetoric and Music: 4. Affects,' *New Grove Dictionary of Music and Musicians*, 2nd ed. [online resource], <www.grovemusic.com> (accessed 1 May 2005).