# Modeling Perceived Emotion With Continuous Musical Features

*Emery Schubert*. **Music Perception**. Berkeley: Jun 2004. Vol. 21, Iss. 4; pg. 561

**Abstract (Summary)**

The relationship between musical features and perceived emotion was investigated by using continuous response methodology and time-series analysis. Sixty-seven participants responded to four pieces of Romantic music expressing different emotions. Responses were sampled once per second on a two-dimensional emotion space (happy-sad valence and aroused-sleepy). Musical feature variables of loudness, tempo, melodic contour, texture, and spectral centroid (related to perceived timbral sharpness) were coded. Musical feature variables were differenced and used as predictors in two univariate linear regression models of valence and arousal for each of the four pieces. Further adjustments were made to the models to correct for serial correlation. The models explained from 33% to 73% of variation in univariate perceived emotion. Changes in loudness and tempo were associated positively with changes in arousal, but loudness was dominant. Melodic contour varied positively with valence, though this finding was not conclusive. Texture and spectral centroid did not produce consistent predictions. This methodology facilitates a more ecologically valid investigation of emotion in music and, importantly in the present study, enabled the approximate identification of the lag between musical features and perceived emotion. Responses were made 1 to 3 s after a change in the causal musical event, with sudden changes in loudness producing response lags from zero (nearly instantaneous) to 1 s. Other findings, interactions, and ramifications of the methodology are also discussed.

**Full Text** (10699 words)



Music PerceptionSummer 2004, Vol. 21, No. 4, 561585The relationship between musical features and perceived emotion was

investigated by using continuous response methodology and time-series

analysis. Sixty-seven participants responded to four pieces of Romantic

music expressing different emotions. Responses were sampled once per

second on a two-dimensional emotion space (happy-sad valence and

aroused-sleepy). Musical feature variables of loudness, tempo, melodic

contour, texture, and spectral centroid (related to perceived timbral

sharpness) were coded. Musical feature variables were differenced and

used as predictors in two univariate linear regression models of valence

and arousal for each of the four pieces. Further adjustments were made

to the models to correct for serial correlation. The models explained

from 33% to 73% of variation in univariate perceived emotion.

Changes in loudness and tempo were associated positively with changes in arousal, but loudness was dominant. Melodic contour varied positively with valence, though this finding was not conclusive. Texture and

spectral centroid did not produce consistent predictions. This methodology facilitates a more ecologically valid investigation of emotion in

music and, importantly in the present study, enabled the approximate

identification of the lag between musical features and perceived emotion. Responses were made 1 to 3 s after a change in the causal musical

event, with sudden changes in loudness producing response lags from

zero (nearly instantaneous) to 1 s. Other findings, interactions, and ramifications of the methodology are also discussed.Received January 26, 2002, accepted October 14, 2003THE power of music to evoke and represent emotions is arguably its

most attractive and, at the same time, most perplexing quality. It is

agreed that the effect is a function of many parameters (Gabrielsson &

Lindstrm, 2001; Juslin, 2000; Scherer, 1991), including the mood, physiological state, cultural background, and preferences of the listener; theAddress correspondence to Emery Schubert, School of Music and Music Education,

University of New South Wales, Sydney NSW 2052, Australia. (e-mail: E.Schubert@

unsw.edu.au)ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions,

University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.561 2004 BY THE REGENTS OF THE UNIVERSITY OF CALIFORNIAALL RIGHTS RESERVED.Modeling Perceived Emotion With Continuous

Musical FeaturesE M E R Y S C H U B E R TUniversity of New South Wales562Emery Schubertperformance style (Sloboda & Lehmann, 2001); the musical style; and the

way in which musical features are combined by the composer

(Gabrielsson & Lindstrm, 2001). Based on the frequently supported

assertion that there is a significant degree of agreement regarding the emotion expressed by a single piece of music, it stands to reason that a nontrivial proportion of emotion in music can be attributed to musical structure, or more specifically, to combinations of musical features. In fact,

within this literature, an assumption is implied that there exists a causal,

underlying relationship between musical features and emotional response.

This assumption is a prerequisite of the present research. Hevner (1935,

1936, 1937) produced the first comprehensive series of publications that

attempted a systematic explanation of the relationship between musical

features and perceived emotion. Since then, a fairly regular stream of publications have attempted to clarify this relationship (for a review, see

Gabrielsson & Lindstrm, 2001). The stimuli used to examine these relationships can be classified along

a spectrum with two extremes: atomistic and ecologically valid. Atomistic

stimuli refer to short auditory items specially produced for an experiment.

They are selected because they afford the researcher with maximal control over the musical parameters. A classic example is the Heinlein (1928) study, in which pairs of chords were presented to listeners who provided an emotional response to them. The pairs consisted of mode manipulations (major-major, minor-minor, major-minor, and minor-major) and loudness manipulations (loud and soft) in all 12 keys. The main problem with these kinds of stimuli is that they lack a musical context. In real-life listening situations, listeners are constantly provided with such a musical context: the moments of music leading up to the present moment of the music. In response to this problem, more and more researchers have used real or ecologically valid musical stimuli to investigate the issue of the effect of musical features on perceived emotion.However, the need for more valid stimuli has necessitated complications in experimental design. Because a typical musical stimulus consists of complex combinations and interactions of loudness, melodic contour, harmony, tempo, durations, texture, timbre, and instrumentation, it is difficult to know which musical feature (or which combination of musical features) could be contributing to the perceived emotion. Two common approaches have been used to investigate this problem. The most common approach is to have the musical features rated by listeners. The raters could be the same people who provide the emotional responses or a separate group of people, usually musical experts, who make purely musical judgments (e.g., Thayer, 1986). With technological advances it has become easier to analyze many of the parameters of musical performances. This has resulted in a second, more sophisticated approach. A musi-Continuous Emotion 563cian can be asked to try to express different emotions in different performances of the same piece of music, and the recording can then be analyzed for concomitant changes in salient musical features (Juslin, 2000; Sloboda & Lehmann, 2001).The methods that use more ecologically valid stimuli have still been unable to provide sufficient information about the nuances in moment to moment changes in both musical features and perceived emotion, with much of the repertoire studied being drawn from the Romantic period of Western art music. Such music is characterized, among other things, by changing emotions. Consequently, to better understand music, or at least Romantic music, it makes sense to track moment to moment changes in perceived emotion. Since the 1980s, an increasing literature has applied to the listening task a continuous response methodology (Schubert, 2001).

Continuous response devices are usually computer-controlled sampling

devices that collect responses at regular, relatively short intervals. Early

continuous response studies tended to record a single emotional dimension that was quite broadly defined, such as emotionality (Sloboda,

Lehmann, & Parncutt, 1997) and tension (Fredrickson, 1995, 1997,

1999; Nielsen, 1987; Krumhansl, 1996), with participants being asked to

make a response on a scale continually while listening to the music. The

response interface might be the mouse on a computer. The mouse movements are then periodically and frequently recorded and synchronized

with the unfolding music. Although this approach helps resolve the problem of tracking moment to moment perceived emotions, it is open to criticism because it lends itself to overly broad, and perhaps not terribly

meaningful, definitions of emotion, which is a complex, multidimensional construct (Schmidt, 1996). A solution to the problem of measuring emotion continuously in time

but meaningfully was proposed by Schubert (1996), who tracked two

dimensions simultaneously instead of one. Based on the theory that emotions can be reduced to two or three important dimensions (Russell,

1989), Schubert selected valence and arousal as bipolar scales aligned at

right angles to form the two-dimensional emotion space (2DES).

Although several such designs are available (Madsen, 1998; Tyler, 1996)

and appear to be fairly reliable (Schubert, 1999), no published study has

yet used such an instrument specifically to analyze how musical features

can be used to predict perceived emotion with this continuous response

methodology.This article reports an experiment and analyses that use continuous

response methodology and time-series analytic techniques to investigate

the relationship between musical/psychoacoustic features and emotions.

The aim is to provide an alternative approach to the understanding of the

relationship between musical features and perceived emotions, and a step564Emery SchubertPARTICIPANTSSixty-seven volunteers participated in the study, with a wide range of ages (M = 30.6,

SD = 12.3). Thirty-eight were females and 29 were males, 60% had played a musical

instrument for more than 10 years, 13% had never played an instrument, 45% reported

that they listened to a great deal of instrumental art music, and 80% expressed a liking for

Western art music.MATERIALThe 2DES (Schubert, 1996) software used in this study was installed on a Power

Macintosh 6300 with a CD drive and a set of headphones connected directly to the computer. The instrument consists of a square with an x- and y-axis dividing the square into

four equal quadrants (Figure 1). The x-axis has a line drawing of a sad face to its left and a happy face to its right. Moving the cursor to the left or right via the computers mousetoward the quantification of emotional response as a function of musical

features by using regression modeling. This study is therefore exploratory

and quasi-experimental.MethodFig. 1. Two-dimensional emotion space. Continuous Emotion 565changes the amount of valence on the emotion scale. A current position box at the bottom left-hand corner of the screen provides an alternative view of the mouses position on

the axis. It indicates both a verbal description of the amount of valence currently selected

(e.g., very positive valence, negative valence) and a 201-point scale ranging from 100%

to +100%. Above the y-axis an aroused label was placed and below a sleepy label.

The current region box provided an alternative readout of the y-axis position using

descriptions and scales similar to that of the valence dimension, but on the line below.

When the music commenced, the 2DES software would begin recording the coordinates

once each second until the music stopped. The musical examples were played over headphones connected directly into the computer. Volume was set to a typical comfortable listening level and was not altered.STIMULIComposers (and performers) do not vary every musical/acoustic parameter within a

piece of music. For example, a movement may be generally fast with some fluctuations in

tempo. However, this movement does not encapsulate the full range of possible tempi,

which appears across a wider gamut of music. If is therefore logical to assume that any

attempt to model emotion as a function of musical features of a single piece will be limited. We would not expect the same equation for every piece because not all features will

be varied across their full range (in addition to the other complications raised by Scherer

& Zentner, 2001). To create a piece that includes such variation might violate the ecological validity conditions that this study aims to maintain. Some composers make it an aesthetic goal for a symphony to contain the world (paraphrase of Mahler by Bonds, 2003),

but the artistic output of such composers tends to be long works that were not considered

suitable for the present experiment. An alternative approach is to select a range of stimuli

that together could capture a reasonably wide range of feature variations. Four pieces of

music were drawn from the Romantic repertoire and were selected to reflect a variety of

possible moods expressed. All selections were taken from the CD Discover Classical Music

(1993). The pieces chosen (with their abbreviations in parentheses, and a short description

justifying their selection) were as follows: Slavonic Dance No. 1 Op. 46 by Antonin Dvorak (Slavonic Dance)

This piece is mostly in a major mode and is in a fast, furiant (Tyrrell, 2004)

tempo, with an accelerando at the end. There are numerous changes in loudness, both sudden and gradual. Overall, the piece seems to evoke highly positive and arousing emotions. It was speculated that it would therefore occupy the top right-hand quadrant of the defined emotion space (Figure 1). Morning from Peer Gynt by Edvard Grieg (Morning)

A slow piece, with gentle, undulating motifs. A major crescendo characterizes the climax of the piece, but the mood of the piece is essentially peaceful and serene. These characteristics mean that it should be located in the bottom right quadrant of the defined emotion space. Adagio movement from Joaquin Rodrigos Concierto de Aranjuez (Aranjuez)

Relative to the other stimuli, this is a long movement (>10 minutes) encapsulating numerous, though mostly negative, moods. The opening theme

played by the oboe and then the solo guitar, seems to express sad, if not

painful, emotions, which become more intense during the tutti orchestral

versions. Some intense, louder, and more dissonant passages such as the

main guitar cadenza seem to express even greater arousal. The movement is

therefore expected to cover a range of emotions expressible in the two lefthand quadrants of the emotion space. Pizzicato Polka by Johan Strauss Jr. and Josef Strauss (Pizzicato)

This is a short, lighthearted, almost comical piece, played entirely pizzicato.

Toward the end, the playing becomes louder and faster, thus suggesting an

increase in arousal, but the piece always appears to express positive emo-566Emery Schuberttions, albeit less extreme than those expressed by Slavonic Dance and

Morning. For these reasons, the piece should occupy the center right of the

emotion space, but mostly the top right quadrant.PROCEDUREParticipants were tested one at a time and completed a questionnaire and some training. The training was to ensure that they understood what was meant by the two dimensions and to ensure that they became competent at moving the mouse around the screen.

In the training stages, participants used the 2DES in its asynchronous (postperformance)

mode to rate emotion expressed by words and pictures of facial expressions. (The findings

of responses to words and pictures of faces using the 2DES are reported in Schubert,

1999.)In the main experiment, one of the four pieces was selected at random. The participant

was asked to judge the emotion the music was trying to express, and not the emotional

response they may be feeling. This important constraint was requested in response to literature suggesting that it is easier to agree on emotion expressed by music than the emotion evoked in the listener (Campbell, 1942; Francs, 1958/1988, pp. 244245; Hampton,

1945; Swanwick, 1973).At the end of each piece, participants completed some additional questions (not reported here) and were given an opportunity to rest before commencing the next randomly

selected piece. Responses were stored on computer files for subsequent analysis. Sound

files of the recordings were analyzed to produce the musical feature variables.ANALYSIS OF MUSICAL FEATURESMethods of coding musical features have varied widely. Important and constructive criteria for coding musical features in the present study are that they should be perceptually

relevant and objectively quantifiable. For example, it makes sense to code loudness as a

continuum from soft to loud because listeners are able to perceive and distinguish different gradations of loudness with reasonable agreement. Consequently, by selecting musical

features where such perceptual gradations occur, they are likely to obtain data that index

an objectively quantifiable aspect of music. For the purpose of the present study, analytic

techniques that could deal with interval-scale variables, such as correlation and regression modeling (LHommedieu, 1992), were employed. This ruled out coding involving

dichotomization of potentially useful information.Only five musical features were selected initially. The limitation was made because too

many musical features may increase the chance of falsely detecting a relationship (Type I

error), thus threatening the validity of the regression model.Melodic contour, tempo, loudness, texture, and timbral sharpness were selected for

analysis because they could be coded as interval-scale variables, and together produce a

reasonable picture of the musical signal. Self-reported emotional responses to melodic contour, tempo, and loudness have been investigated by past researchers. Exploratory variables were formulated from the last two features in the form of the number of instruments

playing (for texture) and the centroid of the frequency spectrum (for timbral sharpness).

They were chosen because they complemented the other three features and could be coded

as interval-scale variables. These are explained further in the following sections.Melodic ContourIn the music under investigation, melody may be defined as a perceptually prominent

and identifiable sequence of pitches, extractable from a multivoiced musical texture. Put

simply, melodic pitch is the coded stream of pitches that represent the melody in a given

piece of music. Melodic contour is defined here as the note to note pitch changes in a

melody and can be represented as the difference transformation of the melodic pitch stream.Continuous Emotion 567(1)(2)A variety of techniques can be used for coding pitch (see Lloyd, 1980; Rossing, 1990). Each

pitch in a melody can be coded according to pitch category or semitone count (Lloyd, 1980,p. 789). Using a rising, equal temperament (enharmonic equivalent) chromatic scale, each

pitch is represented by a consecutive note number. The note number is calculated by combining the register number, indicating the octave from 0 to 9, and the chroma number, indicating

the pitch class from 0 to 11 (C = 0 and B = 11). This procedure can be summarized as:Pitch Category = 12 Register + ChromaFor example, the pitch of middle C (C4), which is in register 4 and has a chroma of 0, is

evaluated as:12 4 + 0 = 48Such a system is used for coding pitch in MIDI devices, where the units of coding are

note number. This notation is adopted even though no MIDI devices were used for coding, because the music consisted of recordings of acoustic performances. The MIDI note

number system is a standardized, objective way of coding pitch and it is perceptually valid.

A higher number corresponds to a higher perceived pitch.An example of how melodic pitchand therefore contourwas coded is shown in

Figure 2. This series was sampled once per second as shown at the bottom of the figure. If

more than one melodic pitch occurred within a 0.5-s window on either side of the sample,

the note with the most extreme value (turning point) was coded. This refers to a global

turning point rather than a local turning point within the sample (i.e., turning point with

respect to adjacent samples not already selected). If there were no global turning points,

the pitch closest to the sample was selected. For example, a sample may appear to have an extreme note with respect to the sample window that is different from that of the actual

point where the melody reaches a turning point (an example appears in Figure 2 at t4,

described later). This has the effect of filtering out ornaments and producing the skeleton

of the melodic line, and it requires a simple differencing transformation to code melodic

contour. This is a useful technique for coding melodic contour, which is typically coded

with far less resolutionthat is, having just three values: up, down, and the same

(Edgeworthy, 1985; Kim, Chai, Garcia, & Vercoe, 2000). Consider the example in Figure2. The first note, B, is the most extreme note for the first sample (at 0 seconds, or t0). The

E (not the F) is coded at t1 because it is a turning point. G is coded at t2 because of its

proximity to the sample (there is no clear extreme in this window). At t3 the B at the beginning of the bar is used because it is a melodic extreme, and at t4 the E is coded. For a similar method of coding melodic contour, see Zhu and Kankanhalli (2002).Fig. 2. Pitch coding sample: opening melody in Morning. Demonstrates melodic coding

and how contour is affected by sampling at a rate of once per second. The effect is that it

smoothes out rapid ornaments from the melodic line. The example shows sampling of the

local global turning point, discussed in the text.568Emery SchubertThe predictions of the emotional effect of melodic contour are still tentative and contradictory, making this an interesting variable to investigate. Gerardi and Gerken (1995)

found a weak relationship between happiness and rising contour (instead, modality was

the strongest predictor of emotional valence). And even this weak finding is distinct from

Cookes (1959) theory, which proposes that rising contour represents outgoing emotions

rather than happy ones. The confusion is further highlighted by reports of some affect of

melodic contour on the arousal dimension of emotion (Peterson, 1994; Scherer &

Oshinsky, 1977) instead of valence as suggested by Gerardi and Gerken.TempoTempo refers to the rate at which the underlying beat or pulse of music progresses. It

is easily quantifiable and often reported in units of beats per minute. The number of beats

played per minute provides an objective measure of musical tempo. Several music-emotion

studies have dichotomized tempo into slow and fast. However, the beats per minute (bpm)

measure provides a far richer and more definitive source of tempo coding.To code tempo, each recording was transferred to a computer-readable sound file and

opened by sound-editing software, SoundEdit 16 (1994). The software enabled visual representation of the musical signal. Available for display were the waveforms and spectra of

the sound (see Figure 3). Tags were placed at the beginning of each bar (measure) of the

display so that the duration of each bar could be measured. These values were then tabulated and converted into bpm, according to the formula:bpm = 60 beats / (durationn + 1 durationn)where beats is the number of beats in a bar (measure), duration is the number of seconds

elapsed from the beginning of the piece until a point in time denoted by the subscript: The onset of the first beat of the current bar is denoted by subscript n and the onset of the following bar is denoted by subscript n + 1. Finally, these values were matched with the sound

recording on a second by second basis, producing a tempo code for any time, t seconds,

corresponding to the tempo of the bar (measure) playing at that time.As an example, a section of the Slavonic Dance is shown in Figure 3. The waveform

and spectogram for the entire piece is shown at the top, and then section by section magnifications until a few bars of the piece appeared on the computer display. The reduced

score below the waveforms and spectra indicates the corresponding point in the music for

the last waveform and spectrum. The loud section, articulated by minim cymbal crashes,

provides an unmistakable hemiola: The cymbal crashes give rise to the sharp peaks in the

frequency spectrum. Subsequently, the frequency spectrum provides a convenient indicator of the bar onset. However, the sudden change to the soft section makes the onset of

the following bar difficult to ascertain. Only by playing the section several times could a

rough estimate be made of the bar onset (for a review of two automated approaches and

algorithms, see Desain, 1993). The existing literature generally supports the idea that

tempo is positively related to the arousal dimension of emotion (Gabrielsson &

Lindstrm, 2001).LoudnessTrue measures of loudness are involved and complex (Campbell & Greated, 1987;

Moore & Glasberg, 1996; Moore, Glasberg, & Baer, 1997; Rossing, 1990; Zwicker &

Scharf, 1965). However a reasonably adequate approximation can be obtained by calculating the A-weighted decibels (dBA) of the signal. A-weighting transforms amplitude

information to account for some of the nonlinearities in the human perception of amplitude. It models the greater sensitivity of the ear to the frequency range found in speech,

and the reduced sensitivity at lower and very high frequencies, as shown in Figure 4.

Loudness calculations were made for each second of music by using PsySound software

(Cabrera, 1997, 1999). As with tempo, loudness appears to be positively related to arousal(3)Continuous Emotion 569Fig. 3. Sample waveforms and spectra of Slavonic Dance. SoundEdit 16 (1994) screen

dumps. Dotted arrows indicate range of blow-ups. Top row indicates time elapsed in seconds, second row indicates bar onsets (cue number). Cue 9 to 10 covers two bars of music.

The first screen dump is of the entire piece. Condensed score of Cue 9 to 10 is shown

below the screen dumps.570Fig. 4. A-weighting curve. Adapted from Campbell and Greated (1987, p. 132).Emery Schubert(Gabrielsson & Lindstrm, 2001). However, the relative contribution of loudness and

tempo has received little attention.TextureMusical texture provides an opportunity to code a higher level musical feature and to

investigate a parameter that has been previously neglected in emotion-music research

(Bruner, 1990; Levy, 1982). Texture may simply and objectively be coded as the number

of voices sounding at any given time, as according to the musical score. A similar approach was used by Nielsen (1987), who referred to this variable as instrumentation. Nielsen (1987) also referred to the thinness of texture, no doubt in the sense of Meyer (1956) in that a thin texture should lead to an increase in tension. This was more a comment in passing than a full-fledged thesis. No other references to texture were cited that used a coding strategy similar to the strategy I am proposing. The coding procedure used here involved simply counting up the number of independent voices at a given point in the score. The decision as to whether or not to count a voice was largely pragmatic, and the following rules were applied:1. An instrument that doubled another instrument of the same section and in the same octave was omitted from the count. For example, two flutes playing in unison were counted as only one. A cello and double bass both sounding C3 were likewise counted as only one because they were both members of the string section. (My use of the term section appears somewhat tenuous here. Celli and double basses are really members of the same group [the chordophone group] rather than the same section. However, the term section helps to distinguish instrumental timbre; for example, theContinuous Emotion 571flute section from the clarinet section, and the trumpet section from the horn section. A violin and a viola playing the same pitch will sound far more homogeneous than these other combinations, and so the distinction I use when referring to a section is necessary.) However, a flute and an oboe playing the same note would be counted as two because they belonged to different instrument families, and the effect would quite likely change the overall texture.2. The same instrument producing two different notes simultaneously counted as two notes. A violin playing double stopped added two to the texture count.3. Sustained notes were added to the count. Excluding doublings, as discussed earlier, any instrument that was playing, as according to the musical score, was added to the texture count.There was no predictive theory cited in the literature regarding texture. The use of this variable and its relationship with emotion is therefore exploratory.CentroidTimbre is a physically and perceptually complex element of music. It does not code neatly into an interval scale that has perceptual validity. Instead, timbre is multidimensional both in terms of its physical features and in terms of its perceptual characteristics (Krumhansl, 1989).Grey (1977) found relationships between particular physical parameters and perceptual judgments of timbre. Three physical dimensions of timbre were found to have various strengths of relationships with perceptions of timbre. Krimphoff, McAdams, and Winsberg (1994, cited by Donnadieu, McAdams & Winsberg, 1994, p. 311) found a high correlation between the centroid of the frequency spectrum and the perceived brightness of the sound (r = 0.94). The labels used to describe the perceptual quality have varied.

Bismark (1974) referred to sharpness, whereas Kendall and Carterette (1996, p. 91) designated the label as nasality. The physical representation of this timbral brightness

correlate (i.e., the A-weighted [explained under the subheading Loudness], frequency

spectrum centroid) was employed, and represented mathematically as follows:C = where Fn is the frequency of partial (or harmonic) number n and An is the amplitude of the

corresponding partial. Equation 4 produces a frequency-spectrum centroid position on the

frequency axis, with the units of measurement being in Hertz.1 The coding of brightness

using frequency spectrum centroid was implemented by using Cabreras software (1997).

The algorithm applied the A-weighted (see Figure 4) frequency spectrum centroid (taken

from Hall, 1987, p. 34), where the weighting is applied to octave bands. This produced a

perceptually more realistic quantification of spectral brightness than the unweighted ver-i Fn An

n = 11. An anonymous reviewer pointed out that spectral centroid also includes information

about pitch height and not just brightness. To control for pitch height, the reviewer recommended dividing the centroid by an average of the notated pitches. To do so for a large

number of orchestral chords, however, raises not the only practical problems of identifying all contributing pitches and performing the calculation, but also the subjective decision

of how to weight the average for the intensity of the contributing pitches or spectral components. Therefore, in the absence of this calculation, brightness as used in the present

context also includes pitch height information.i An

n = 1(4)572Emery Schubertsion. Other techniques for calculating spectral centroid are discussed in Kendall and

Carterette (1996). Although no studies on the effect of centroid on emotional response

have been published, Gabrielsson and Juslin (1996) have reported a trend for bright

sounds to express higher arousal than softer sounds.QUANTITATIVE MODELING PROCEDUREA linear regression type model using the ordinary least square approach was employed

in which the musical feature variables were used as predictors of perceived emotions for

second by second response. Two sets of univariate models are reported, one for arousal

and the other for valence. Traditional linear multiple regression models are not an adequate way to model the data collected in this experiment, mainly because capturing data

from moment to moment violates the assumption of independence of data points (see

Ostrom, 1990; Schubert & Dunsmuir, 1999). The solution proposed by Schubert and

Dunsmuir was to difference the variables and then model out effects of serial correlation

by using an autoregressive term. Differencing involves subtracting each value in the time

series from its preceding value. This transformation produces a gradient time series, indicating the amount of change in the variable (e.g., second to second change in arousal or change in loudness). In the case of melody, differencing produces melodic contour by definition. Each musical feature variable was duplicated and lagged by 1 s (i.e., delaying it by 1 s

from the original variable). The lagged variable was then lagged again, delaying it by 2 s

from the original variable. The process was repeated until a 4-s lag variable was produced.

Lagging the variables is necessary because it is called for by the assumption that emotional response is causally related to musical features. The assumption infers that an emotional response will occur close to or a short time after the causal musical event. Hence, for

analysis, musical feature variables must be included in a range of delayed (lagged) forms

before regression modeling. The choice of a 4-s lag was based on preliminary exploration

and on other continuous response literature (Krumhansl, 1996; Schubert, 2001; Sloboda

& Lehman, 2001).An autoregressive term was added to the traditional linear regression equation. The

term consists of a linear combination of previous error terms and is labeled AR1. This term

factors in a kind of memory that is implicit when dealing with time-series data. With conventional experimental designs, the assumption of independence might be satisfied by segmenting the music into small blocks, and testing each one separately, but this destroys the

musical context. Autoregressive adjustment allows real pieces to be analyzed in a more

realistic way because the effect of musical context or the emotional memory can be

quantitatively modeled out. For example, a large autoregressive coefficient suggests that

the response to a piece at a given point is determined largely by what went on before,

rather than being determined by the actual musical feature changes occurring near that

point in time (Gottman, 1981).In the present study, by expressing this error term as a function of only the previous

sample error generated by the model, the assumptions of the model were satisfied: That is,

the final models displayed no serial correlation in their residuals (Ostrom, 1990). Results and DiscussionPERCEIVED EMOTION WITH EACH PIECESlavonic DanceThe final model for the arousal gradient response to Slavonic Dance is

shown in Table 1. All coefficients have a positive sign. This means thatContinuous Emotion 573TABLE 1First-Order Autoregression Models Summary TableLagLagMF 0 1 2 3 4Arousal ModelsSlavonic Dance

R2 = 0.73 AR1 = 0.49 N = 225Tempo .02 .01 .03 .05 .04

Centroid .12** .30** .17** .06 .07

Loudness .52** .47** .24** .25** .13**

Melody .03 .03 .04 .00 .05Texture .04 .00 .05 .04 .04Aranjuez

R2 = 0.58 AR1 = 0.66 N = 654Tempo .01 .01 .09** .07** .06

Centroid .07 .07 .03 .00 .03

Loudness .07** .16** .13** .12** .05

Melody .02 .07** .06 .01 .06

Texture .04 -.04 .14** .14** .06**Pizzicato

R2 = 0.35 AR1 = 0.44 N = 150Tempo .11 .05 .01 .12** .11

Centroid .07 .05 .06 .05 .06

Loudness .26** .22** .17** .13 .09

Melody .07 .08 .02 .02 .04

Texture .08 .30 -.02 .02 .02Morning

R2 = 0.67 AR1 = 0.51 N = 216Tempo .06 .07 .03 .16** .08

Centroid .04 .02 .09** .02 .12**

Loudness .06 .73** .70** .99** .42**

Melody .02 .03 .03 .06 .04Texture .12** .02 .02 .01 .10NOTEStandardized coefficients are shown to allow comparison across variables. MF = musical

feature, R2 = approximate model fit, AR1 = coefficient for first order autoregression term, N = number

of samples (equivalent to the duration of the stimulus in seconds), Tempo = coefficient for change in

tempo (units of variable: beats per minute), Centroid = coefficient for change in spectral centroid (units

of variable: Hertz), Loudness = coefficient for change in loudness (units of variable: A-weighted decibels), Melody = coefficient for melodic contour (units of variable: MIDI note number), Texture = coefficient for change in texture (units of variable: number of instruments playing). **p < .01.MF 0 1 2 3 4Valence ModelsSlavonic Dance

R2 = 0.62 AR1 = 0.50 N = 225Tempo .08 .13** .09** .07** .05

Centroid .06 .05 .08 .04 .05

Loudness .32** .26** .01 .04 .01

Melody .01 .04 .04 .04 .01

Texture .02 .04 .03 .07 .04Aranjuez

R2 = 0.33 AR1 = 0.48 N = 654Tempo .01 .00 .07 .04 .02

Centroid .04 .03 .07 .06 .05

Loudness .01 .04 .02 .02 .01

Melody .04 .19** .39** .38** .22**

Texture .03 .01 .08 -.01 .06Pizzicato

R2 = 0.38 AR1 = 0.50 N = 150Tempo .02 .10 .06 .06 .01

Centroid .02 .01 .04 .02 .14

Loudness .14 .33** .19** .20** .10

Melody .03 .07 .08 .15 .01

Texture .00 .02 .13 .04 .03Morning

R2 = 0.40 AR1 = 0.40 N = 216Tempo .06 .10 .02 .10 .03

Centroid .07 .00 .03 .02 .12

Loudness .05 .10 .09 .09 .07

Melody .01 .00 .02 .11 .08

Texture .01 .14** .29** .36** .29**changes in the arousal will be in the same direction as the change in each

musical feature when other musical feature variables are held constant. In

other words, the partial correlations between musical feature and perceived emotion are positive. Approximately 73% of the variance was

explained by this model. This good fit, and examination of the coefficients, suggests that, in line with previous literature, the model successfully explains changing arousal response in terms of loudness gradient (see

Gabrielsson & Lindstrm, 2001), centroid gradient (more or less in agreement with Gabrielsson & Juslin, 1996), and a first-order autoregressive

process. Increases in loudness and brighter timbre each play a part in

increasing arousal response gradient.574Emery SchubertThe standardized coefficients indicated in Table 1 show the relative

contribution of each variable to changes in arousal. The peak change in

arousal could be traced to the first second after the change in centroid.

However, changes in loudness occurred within 1 s of arousal change so as

to produce a large coefficient (0.52) instantaneously (i.e., with zero lag).

The short lag can be explained by the sudden changes in loudness that

occur in this piece. Schubert and Dunsmuir (1999) argued that a lag of

around 3 s was typical of the emotional response time after a particular

musical feature changed, but they also pointed out that the system

changes when the loudness changes suddenly. Slavonic Dance contains

several sudden changes in loudness, most notably the opening, loud

chord. In such cases, response times become noticeably shorter. This

explains the high coefficients for changing loudness at short lags.The centroid relationship corresponds to the use of higher instruments

to add loudness to the orchestral sound in loud passages. In this respect,

centroid acts as an approximate duplication of loudness. That is, to make

the orchestra sound louder, higher instruments, such as trumpets, piccolo,

and cymbals are added.This Slavonic Dance Valence model explained approximately 62% of

the variance. A comparison of the valence and arousal blocks for Slavonic

Dance in Table 1 shows how the coefficients for the valence model are

smaller than for the corresponding arousal model, meaning that the relative contribution of these variables is less able to explain changes in valence. One of the reasons for the lower contribution, and the lower

explained variance, may be that there was a missing variable (Ostrom,

1990). In Slavonic Dance, the loud, fast sections often occur in a major

key and there is an accelerando passage that moves from a minor key to

a major key. This is an example of the composer using several musical features together to enhance the portrayal of particular moods. However, it

does appear that increasing tempo or loudness may each independently

contribute to the increasing rate of happiness of the piece. The two features are distinguished by their proportional contribution to valence and

by their lag structure: Loudness dominates tempo, and changes in loudness are associated with shorter latency (01 s before a change in valence),

compared with tempo (13 s).Aranjuez Tempo, loudness, and texture each make a significant contribution to

the perception of arousal according to the Aranjuez arousal model shown

in Table 1. This model explains approximately 58% of the variation in

response. Melodic contour makes a marginal contribution. A considerable

literature supports the idea that tempo and loudness contribute to arousal.Continuous Emotion 575However, the present model shows a strong relative contribution coming

from loudness, with significant changes in arousal occurring 1, 2, and 3 s

after the change in loudness. Given that both tempo and loudness varied

considerably in this movement, it seems that loudness might be a stronger

contributor to arousal than tempo. The faster arousal response to changes

in loudness (0.16 coefficient at lag 1) also supports the dominance of

loudness in producing arousal response.The more unusual finding here, however, is that texture makes a strong

contribution to arousal. No literature has been cited that predicts this

result. An examination of the texture reveals that when the full orchestra

is playing it often sounds louder, such as the first tutti entry after the oboe

and guitar solos at the opening of the movement. Hence there is a possible confounding of texture with loudness. However, there are passages

when texture receives a relatively low count, while arousal remains high,

such as the long guitar cadenza. Further, changes in texture take longer to

register their influence upon arousal compared to lags for other variables,

with lags in the range of 2 to 4 seconds (Table 1). This result suggests that

texture may be making some independent contribution to the perception

of arousal in the music. Melodic contour has a marginal effect, and only at one lag (1, with standardized coefficient of 0.07). As described later, Aranjuez has a long, arch-shaped melody. As the melody rises, the playing tends to get louder. As the phrase comes to its downward end, the playing gets softer. Again, it is possible that melody is confounded by loudness.The model describing the valence gradient response to Aranjuez (Table1) explained 33% of the variance. It reveals a relationship between melodic contour and valence. Aranjuez is the only stimulus that contained a long melodic phrase structure. The first melodic subject takes nearly 40 s to expose. It is an arch-shaped melody, slowly rising, and then falling. The first subject is played twice, on the oboe and then on the solo guitar. The second subject is related to the first and takes around 25 s to expose. It consists of a downward melodic shape. This melodic structure provides a good opportunity to examine the relationship between emotional response and melodic movement. The finding is in agreement with the research of Gerardi and Gerken (1995), which posits that melodic contour is positively related to valence. The present finding adds to this by identifying a valence response lag of between one and four seconds, typically two or three seconds after the change in melodic contour.An important reason for the smaller explained variance relative to the other models is that only one variable, with four statistically significant lags, appears in the equation (Table 1). The stepwise regression process determined that these four lags of melodic contour significantly added to the fit of the model, while possessing significantly nonzero coefficients in the AR(1) model. The result suggests that something else, apart from576Emery Schubertloudness, texture, centroid, and tempo, is required to further explain the bulk of the response variance. Articulation may have been such a predictor. The middle section of the movement consists of a short, light-hearted, scherzandesque musical idea. Here valence increases, but there was no variable coded to identify this. Given the importance of articulation in affecting musical character (Fabian & Schubert, 2004), this is a variable that should be considered in future research. Harmony is also likely to have contributed to valence, as the amount of consonance and dissonance varied greatly in this piece: For example, the long cadenza employed some highly complex chords.PizzicatoArousal gradient in Pizzicato was best modeled in terms of lagged loudness variables and a lagged tempo variable, as summarized in Table 1. The model explained about 36% of the variation in changing arousal. Few of the investigated musical parameters change in this piece, though notable

is the sudden increase in loudness in the second part of the trio section.

This section consists of eight bars repeated, and connects the trio to the

da capo. It is the first loud section of the piece. Further, the loudness

begins quite suddenly. This is associated with a faster increase (smaller

lag) in arousal response. The opening section, shown in Figure 5, indicates

a section with sudden changes in loudness, which also explains the high

loudness coefficient at the shorter lags (Table 1). Together, these contribute to the relatively high coefficient for loudness (0.26) at lag zero. The

piece shifts in tempo enough to produce some effect upon arousal, but as

with other arousal models, loudness has the dominating influence even

though the musical forces are much smaller in Pizzicatoa string orchestra with triangle percussion.An autoregressive model for valence gradient in Pizzicato, shown in

Table 1, explained approximately 38% of the variance in changing

valence. It predicts a positive relation between loudness and valencefor

example, when rate of loudness decreases, so does rate of valence. Musical

features other than loudness provide clues to this finding. The first four

bars consist of short-duration notes in a major mode, with loud and soft

chords juxtaposed and with rising pitch. The silent pause occurs where

there is an expectation of another loud chord (Figure 5). This combination of musical features provides ingredients appropriate for the expression of humor (Mull, 1949) and hence a rise in valence.MorningThe autoregressive model for changing arousal in response to Morning

explains approximately 67% of the variation in changing arousal.Continuous Emotion 577Fig. 5. Juxtaposition of loud and soft in Pizzicato Polka. Piano score reduction from

Strauss and Strauss (1943, p. 16).Morning consists of numerous gradual fluctuations in loudness, and these

are strongly aligned with arousal response. As shown in Table 1, arousal

varies significantly at four lags after a change in loudness, with a very

strong peak at the third lag (coefficient of 0.7), although the instantaneous

(zero) lag found in other arousal models is absent. This may be attributed

to the absence of sudden changes from very soft to very loud in this movement. According to musical scores of the piece, tempo does not vary greatly in Morning, but there exists a general pattern in the performance: The

orchestra speeds up from an opening tempo of around 48 bpm to about

55 bpm as the middle section of the piece is reached, and then slows down

to around 40 bpm near the quiet ending. Despite this, the model suggests

that loudness still dominates over tempo in determining arousal response. An unusual finding for the Morning arousal model is the statistically

significant negative correlation with centroid and texture (Table 1).

Reduction in texture is associated with slight, but sudden (zero lag)

increase in arousal. Texture roughly follows the outline of arousal at the

beginning and middle of the piece. However, toward the end, as the mood

of the piece softens, there are points where Greig uses a large number of

players playing either softly, or playing with a slight crescendo, as in the

very last chord. It seems that these occasions are sufficient to cause the

negative correlation. The negative centroid correlation can be explained

by the addition of the lower brass to increase the volume. This leads to

increase in arousal, but the centroid has actually decreased. Examples of

this occur in the second theme, discussed next. At the end of the theme,

arousal increases as does loudness, but at the same time, lower pitched

instruments are added. Hence, centroid moves down as the loudness

increases. This creates some complications regarding the causal contribution of centroid to arousal, for it could be that the variable happens to be

fluctuating in the opposite direction to loudness. The Morning valence model explains approximately 40% of the

response variance and significant contributors were entirely restricted to578Emery Schubertlagged texture variables, with the strongest change in valence occurring

approximately 3 s after a change in texture, as shown in Table 1.

Noticeable and frequent changes in texture occur in the second theme,

which commences with a melody outlining a major third (third, fourth,

and fifth degrees of the minor chord harmony) using a small number of

instruments. This is followed by a buildup of instrumental forces ending

at the cadence (closure point) in a new key. This section demonstrates how

harmony and texture are being used together to produce increased

valence. However, it also suggests that texture is not acting independently to produce the arousal response. LAG STRUCTUREIn the present study, lag structure refers to the delay between change in

a musical feature and the perceived emotion that follows. The justdescribed results demonstrate a logical unfolding of emotion after a musical event: Quite often the perceived emotion is correlated at a lag close in

time to the musical feature, with the coefficient rising and then falling

away as the lag increases. An event that consistently follows another in

time presents strong evidence of a causal relationship. In general, there is

a 1- to 3-s lag, consistent with Krumhansls (1996) and Sloboda and

Lehmanns (2001) reports, with a peak occurring 2 to 3 s after the emotion inducing musical agent. Shorter lags (01 s) are characterized by

the effect of sudden changes of loudness upon arousal.Making assertions about lag structure requires consideration of both

the mathematical and psychological processes involved in producing the

lag. The present analysis is based on a linear model. However, responses

will not always occur a linear distance (in time) after an appropriate

musical feature changes. Schubert and Dunsmuir (1999) identified these

non-linearities by examining and categorizing outliers from a linear

model. One of the findings of their study was that sudden changes in a

musical feature reduce the delay in response, as found in the present study.

Psychologically, reaction time can be represented as:t(auditory perception) + t(cognitive processing [including

task decision]) + t(physical response)where the first two terms represent the perceptual and cognitive processing time and the last term corresponds to the time taken in the actual

process of moving the mouse on the emotion space. It can be shown that

this last term occupies a relatively small amount of time in the total lag.

As Schubert and Dunsmuir demonstrated, when large changes in response

were required because of, for example, sudden changes in loudness,

response times reduced from a typical 2- or 3-s lag to a 0- or 1-s lag. That(5)Continuous Emotion 579is, the larger physical distance traveled by the mouse did not produce an

overall increase in reaction time, but a reduction. If physical response time

made a significant contribution to overall response time (i.e., lag structure), then the need to travel large distances on the emotion space should

have resulted in increased response lag. At other times during the listening period, mouse movements are usually continuous adjustments from

one nearby point to another, and as such, physical response time would

not be very large. Clearly, the cognitive processing time required to perform the task is proportionally greater than the physical movement time.

Also clear from the Schubert and Dunsmuir study is that sudden increases in loudness may even partially bypass the cognitive processing system

because of the activation of a more primal, startle effect, which is an involuntary response to a sudden increase in sound level (Gaston, 1951;

Howard & Ford, 1992). However, this still does not guarantee that the

emotional response time diminished instead of the task response time.

Indeed it is difficult to abstract the two processes in the present design.

The emotional response and the movement task may even be operating in parallel (which is why the two components are grouped together in Equation 5). Until further work on emotional/cognitive processing and reaction time on the 2DES is made, the measuring paradigm cannot be assumed to be able to pinpoint the time before an emotion is felt. Therefore the assumption that this study is reporting perceived emotion per se is necessarily a simplification.Interactions Only a few studies have focused on how musical features interact with one another to produce emotional effects (e.g., see Flowers, 1988; Gerardi & Gerken, 1995; Schellenberg, Krysciak, & Campbell, 2000). Although the present study does not explicitly or statistically analyze interactions, on several occasions interactions seem to be a sensible option for seeking to better understand how emotions could be explained in terms of combinations of musical feature variables. Apart from the finding of a relationship between loudness (and to a lesser extent, tempo) and arousal, a higher level of complexity is required to discover and understand the underlying principles. For example, the articulation, the mode of the piece, and the overall tempo appeared to have a controlling effect on whether changes in either tempo, texture, or loudness affected valence or whether they affect arousal. Minor mode, slow tempo, and legato articulation appear to divert the effect of changing tempo to the arousal dimension and away from the valence dimension. For instance, a possible interaction between mode and loudness may be hypothesized as indicated in Figure 6.580Fig. 6. Hypothesized interaction between tempo and mode.Emery SchubertConclusionsThis study provides an alternative approach to the examination of the relationship between musical features and perceived emotion from traditional postperformance methodologies. It investigated how quantitatively coded musical feature variables could be used to predict perceived emotion in music within a statistically valid framework. Although the equation for emotional response to music is complex (Scherer & Zentner, 2001), the study suggests that a significant proportion of such response could be explained by a handful of musical feature variables. Of particular note here is the finding that more than 60% of variation in arousal response could be explained by musical features such as loudness and tempo. Perceived emotional valence models were not as convincing, a finding that is not surprising given that variables omitted in this study, such as mode and articulation, influence valence (Gabrielsson &

Lindstrm, 2001; Fabian & Schubert, 2004). In this study, these musical

features were not coded because they were not as easy to represent as

interval-scale variables. Instead, the aim in the present study was to examine how well emotional response could be mathematically modeled by

using only a handful of interval-scale musical features. However, the

regression modeling approach does seem promising, and future research

should consider incorporating more variables into the equation, if for no

other reason than to see how much response variation can be explained

by musical feature variables alone.In addition, tentative steps were made to examine how interactions of

musical features might affect emotional response, and how hidden vari-Continuous Emotion 581ables might be identified. Under the assumption that perceived emotion

(or any kind of emotional response to music within a culture) is causally

connected with variations in musical features, exploratory studies such as

the present one seem to be fruitful in gaining greater insights into the

nature of music. The present study also points to a need to develop more

complex models that explain interactions and hidden variables through

further use of continuous response methodology.The autoregressive adjustment of the regression analyses was necessary

to produce valid statistical models. The coefficient was significant for all

models, and a large amount (ranging from 39% to 63%) of the error term

in the previous second could be used to provide a significant increase in

explanatory power of the models. This leaves open the question of how to

interpret this component of the equation. I argue that this component of

the model can be explained as a mathematical representation of a cognitive process in that a listeners emotional response is not just a function of

musical features, predisposition, preferences, personality traits, and a

range of other variables as indicated by Scherer and Zentner (2001), but

also of memory during the listening process. This memory propagates

through the music-emotion system, meaning that every moment in the listening experience is, to some extent, related to every other moment of listening to the same piece. This finding appears not to have been identified

by previous researchers, largely because of the postperformance-analysis

research tradition. Memory during the listening process (and therefore,

musical context) makes an important contribution to perceived emotion,

in addition to numerous other factors such as musical features, preference,

mood, interpersonal stances, attitudes, and personality traits.The results tend to support previous researchfor example, that loudness and tempo affect perceived arousal, and that melodic contour may be

loosely related to valence in some situationsbut in addition help to identify some of the relationships and interactions between musical features.

For example, the study shows a clear dominance of loudness over tempo

in affecting arousal response. The findings also support the use of these

commonly investigated musical feature variables (tempo, loudness, and

melodic contour) because the use of the less commonly used variables

texture and centroiddo not appear to have a clear relationship to perceived emotion. Although this does not demonstrate that centroid and

texture are unrelated to perceived emotion, the positive findings do pose

a caveat to the present research: If the complex, data-driven modeling of

response produces similar results to those of asynchronous, postperformance responses, is there any point in going to the effort of using continuous response methodology? Clearly, the answer is yes, provided there is

a need to do so. The continuous response methodology allows broad

results to be presented, as is the case in the present study, but more importantly, the methodology allows a close-up inspection of smaller sections of582Emery Schubertmusic. This is important because listeners respond to an entire movement,

and so the ecological validity is maintained, as distinct from atomistic,

context-starved research such as the Heinlein (1928) study. The analysis

and modeling technique accounts for the inherent complexities of analyzing time-series data by modeling out the effects of serial correlation. Also,

continuous response methodology enables researchers to understand the

lag structure of emotional response. However, the present study does not

allow the precise pinpointing of feature-to-emotion lag because of the

complication caused by the physical response task. By the same token,

postperformance responses cannot tell us about how long after a musical

event a perceived emotion kicks in.2Bismark, G. von (1974). Sharpness as an attribute of the timbre of steady sounds.Acustica, 30, 159172. Bonds, M. E. (2003). Symphony: 19th century. In L. Macy (Ed.), The New GroveDictionary of Music Online. Retrieved August 29, 2003, from http://www.grovemusic.comBruner, G. C. (1990). Music, mood, and marketing. Journal of Marketing, 54, 94104.

Cabrera, D. (1997). Db&dba&centr.orc [computer software]. Sydney, Australia: Author.

Cabrera, D. (1999). Psysound: A computer program for the psychoacoustical analysis

of music. Mikropolyphonie 5. Retrieved June, 1, 2003, from

http://farben.latrobe.edu.au/mikropol

Campbell, I. G. (1942). Basal emotional patterns expressible in music. American Journal

of Psychology, 55, 17.Campbell, M., & Greated, C. (1987). The musicians guide to acoustics. London: J. M.Dent and Sons. Cooke, D. (1959). The language of music. London: Oxford University Press.

Desain, P. (1993). A connectionist and a traditional AI quantizer, symbolic versus sub-symbolic models of rhythm perception. Contemporary Music Review, 9, 239254.

Discover Classical Music, CD 2 [sound recording] (1993). HNH International Ltd.8.550008-9.Donnadieu, S., McAdams, S. & Winsberg, S. (1994). Context effects in timbre space.Proceedings of the 3th International Conference of Music Perception and Cognition

(pp. 311312). Liege, Belgium: ICMPC.

Edgeworthy, J. (1985). Interval and contour in melody processing. Music Perception, 2,375388.Fabian, D., & Schubert, E. (2004). Expressive devices and perceived musical character in34 performances of Variation 7 from Bachs Goldberg Variations. Musicae Scientiae,

Special Issue 20032004, 4971.

Flowers, P. J. (1988). The effect of teaching and learning experiences, tempo, and mode on

undergraduates and childrens symphonic music preference. Journal of Research in

Music Education, 36, 1934.2. This paper was supported by a University of New South Wales (UNSW) Faculty of

Arts and Social Sciences Postdoctoral Writing Award and the UNSW Vice-Chancellors

Postdoctoral Fellowship. I am deeply grateful to Gary McPherson, Kate Stevens, Joe

Wolfe, Alfonso Egan, and Stephanie Wilson for their assistance, support, and encouragement in the preparation and revision of this manuscript. ReferencesContinuous Emotion 583Francs, R. (1958/1988). The perception of music (W. Jay Dowling, Trans.). Hillsdale, NJ:Lawrence Erlbaum Associates. Fredrickson, W. E. (1995). A comparison of perceived musical tension and aesthetic

response. Psychology of Music, 23, 8187.

Fredrickson, W. E. (1997). Elementary, middle, and high school perceptions of tension in

music. Journal of Research in Music Education, 45, 626635.

Fredrickson, W. E. (1999). Effect of musical performance on perception of tension inGustav Hoists First Suite in E-flat. Journal of Research in Music Education, 47, 4452.

Gabrielsson, A. & Juslin, P. N. (1996). Emotional expression in music performance:Between the performers intention and the listeners experience. Psychology of Music,

24, 6891.Gabrielsson, A., & Lindstrm, E. (2001). The influence of musical structure on emotional expression. In P. N. Juslin & J. A. Sloboda (Eds.), Music and emotion: Theory and

research (pp. 223248). London, England: Oxford University Press.

Gaston, E. T. (1951). Dynamic factors in mood change. Music Educators Journal, 37,4244.Gerardi, G. M., & Gerken, L. (1995). The development of affective responses to modality and melodic contour. Music Perception, 12, 279290.Gottman, J. M. (1981). Time-series analysis: A comprehensive introduction for social scientists. Cambridge: Cambridge University Press.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. Journal of theAcoustical Society of America, 61, 12701277.

Hall, D. E. (1987). Basic acoustics. New York: John Wiley and Sons.

Hampton, P. J. (1945). The emotional element in music. Journal of General Psychology,33, 237250.Heinlein, C. P. (1928). The affective characters of the major and minor modes in music.Journal of Comparative Psychology, 8, 101142.

Hevner, K. (1935). The affective character of the major and minor modes in music.American Journal of Psychology, 47, 103118.

Hevner, K. (1936). Experimental studies of the elements of expression in music. AmericanJournal of Psychology, 48, 246268.Hevner, K. (1937). The affective value of pitch and tempo in music. American Journal ofPsychology, 49, 621630.Howard, R., & Ford, R. (1992). From the jumping Frenchmen of Maine to posttraumatic-stress-disorder: The startle response in neuropsychiatry. Psychological Medicine, 22,

695707.Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance:Relating performance to perception. Journal of Experimental Psychology: Human

Perception & Performance, 26, 17971813.

Kendall, R. A., & Carterette, E. C. (1996). Difference thresholds for timbre related to

spectral centroid. Proceedings of the 4th International Conference of Music Perception

and Cognition (pp. 9195). Montreal, Canada: ICMPC.

Kim, Y. E., Chai, W. Garcia, R. & Vercoe, B. (2000). Analysis of a contour-based representation for melody. Retrieved June 1, 2003, from www.media.mit.edu/~chaiwei

/papers/Kim.pdfKrumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzn andO. Olsson (Eds.), Structure and perception of electroacoustic sound and music (pp.

4353). Amsterdam: Elsevier.

Krumhansl, C. L. (1996). A perceptual analysis of Mozarts Piano Sonata K.282: segmentation, tension, and musical ideas. Music Perception, 13, 401432.

LHommedieu, R. (1992). Regression-based research designs. In R. Colwell (Ed.),Handbook of research on music teaching and learning (pp. 184195). New York:

Schirmer. Levy, J. M. (1982). Texture as a sign in classic and early romantic music. Journal of theAmerican Musicological Society, 35, 482531.584Emery SchubertLloyd, L. S. (1980). Pitch notation. In S. Sadie (Ed.), New Grove Dictionary of Music andMusicians (Vol. 14, pp. 786789). London: Macmillan.

Madsen, C. K. (1998). Emotion versus tension in Haydns symphony No. 104 as measured

by the two-dimensional continuous response digital interface. Journal of Research in

Music Education, 46, 546554.

Meyer, L. B. (1956). Emotion and meaning in music. Chicago: University of Chicago Press.

Moore, B. C. J., & Glasberg, B. R. (1996). A revision of Zwickers loudness model. ActaAcustica, 82, 333345.Moore, B. C. J., Glasberg, B. P., & Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. Journal of the Audio Engineering Society, 45,

224240.Mull, H. (1949). A study of humour in music. American Journal of Psychology, 62,560566.Nielsen, F. V. (1987). Musical tension and related concepts. In T. A. Sebeok & J. Umiker-Seboek (Eds.), The semiotic web 86: An international year-book (pp. 491513). Berlin:

Mouton de Gruyter. Ostrom, C. W. (1990). Time series analysis regression techniques. Newbury Park, CA:Sage.Peterson, M. R. (1994). Musical expressiveness. Unpublished masters thesis, CaliforniaState University, Long Beach. Rossing, T. D. (1990). The science of music (2nd ed.). Reading, MA: Addison-Wesley.

Russell, J. A. (1989). Measures of emotion. In R. Plutchik & H. Kellerman (Eds.),Emotion: Theory research and experience (Vol. 4, pp. 81111). New York: Academic

Press. Schellenberg, E. G., Krysciak, A. M., & Campbell, R. J. (2000). Perceiving emotion in melody: Interactive effects of pitch and rhythm. Music Perception, 28, 155171.

Scherer, K. R. (1991). Emotion expression in speech and music. In J. Sundberg, L. Nord,& R. Carlson (Eds.), Music, language, speech and brain (pp. 146156). Cambridge:

Macmillan. Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. Motivation and Emotion, 1, 331346.

Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules.In P. N. Juslin & J. A. Sloboda (Eds.), Music and emotion: Theory and research (pp.

361392). London, England: Oxford University Press.

Schmidt, C. P. (1996). Research with the continuous response digital interface: A review

with implications for future research. Philosophy of Music Education Review, 4,

2032.Schubert, E., & Dunsmuir, W. (1999). Regression modelling continuous data in music psychology. In S. W. Yi (Ed.), Music, Mind, and Science (pp. 298352). Seoul, Korea: Seoul

National University Press.Schubert, E. (1996). Continuous response to music using a two dimensional emotion

space. Proceedings of the 4th International Conference of Music Perception and

Cognition (pp. 263268). Montreal, Canada: ICMPC..

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two

dimensional emotion space. Australian Journal of Psychology, 51, 154165.

Schubert, E. (2001). Continuous measurement of self-report emotional response to

music. In P. N. Juslin & J. A. Sloboda, (Eds.), Music and emotion: Theory and

research (pp. 393414). London, England: Oxford University Press.

Sloboda, J. A., & Lehmann, A. C. (2001). Tracking performance correlates of changes in

perceived intensity of emotion during different interpretations of a Chopin Piano

Prelude. Music Perception, 19, 87120.

Sloboda, J. A., Lehmann, A. C., & Parncutt, R. (1997). Perceiving intended emotion in

concert-standard performances of Chopins Prelude No. 4 in E-minor. In Proceedings

of the Third Triennial ESCOM Conference (pp. 629634). Uppsala, Sweden: ICMPC.

SoundEdit 16 (Version 1.0) [Computer software]. (1994). San Francisco: Macromedia.Continuous Emotion 585Strauss, J., Jr. & Strauss, J. (1943). Pizzicato Polka [music]. In Echoes of the Ballet forPiano (pp. 1617). Melbourne, Australia: Allan and Co.

Swanwick, K. (1973). Musical cognition and aesthetic response. Psychology of Music, 1,713.Thayer, J. F. (1986). Multiple indicators of affective response to music. Unpublished doctoral dissertation, New York University. Tyler, P. (1996). Developing a two-dimensional continuous response space for emotions

perceived in music. Unpublished doctoral dissertation, Florida State University,

Tallahassee. Tyrrell, J. (2004). Furiant. In L. Macy (Ed.), The New Grove Dictionary of Music Online.Retrieved April 19, 2004, from http://www.grovemusic.com.

Zwicker, E., & Scharf, B. (1965). A model of loudness summation. Psychological Review,72, 26. Zhu, Y., & Kankanhalli, M. (2002). Similarity matching of continuous contours for humming querying of melody databases. LIT Technical Report, February, 2002. Retrieved

June 1, 2003 from www.comp.nus.edu.sg/~cs4241/TR2002.pdf

**References**



**Indexing (document details)**

| | |
|---|---|
| **Author(s):** | Emery Schubert |
| **Publication title:** | Music Perception. Berkeley: Jun 2004. Vol. 21, Iss. 4; pg. 561 |
| **Source type:** | Periodical |
| **ISSN:** | 07307829 |
| **ProQuest document ID:** | 709510651 |
| **Text Word Count** | 10699 |
| **Document URL:** | http://proquest.umi.com/pqdlink?did=709510651&Fmt=3&clientId =11561&RQT=309&VName=PQD |