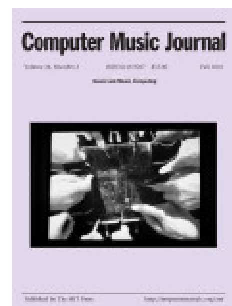
Changing Musical Emotion: A Computational Rule System for Modifying Score and Performance

Steven R. Livingstone
Ralf Muhlberger
Andrew R. Brown
William F. Thompson



Computer Music Journal, Volume 34, Number 1, Spring 2010,
pp. 41-65 (Article)

Published by The MIT Press

Changing Musical Emotion: A Computational Rule System for Modifying Score and Performance

Steven R. Livingstone,*† Ralf Muhlberger,† Andrew R. Brown, and William F. Thompson††**

*Sequence Production Lab
Department of Psychology
McGill University
1205 Dr Penfield Avenue
Montreal, Quebec H3A 1B1 Canada
steven.livingstone@mcgill.ca
†School of Information Technology and Electrical Engineering
University of Queensland
Brisbane, Queensland, 4072 Australia
ralf@itee.uq.edu.au
**Australasia CRC for Interaction Design
PO Box 2008
Kelvin Grove, Queensland
4059 Australia
a.brown@qut.edu.au
††Department of Psychology
Macquarie University
Sydney, New South Wales
2109 Australia
bill.thompson@mq.edu.au

Composers and performers communicate emotional intentions through the control of basic musical features such as pitch, loudness, and articulation. The extent to which emotion can be controlled by software through the systematic manipulation of these features has not been fully examined. To address this, we present CMERS, a Computational Music Emotion Rule System for the real-time control of musical emotion that modifies features at both the score level and the performance level. In Experiment 1, 20 participants continuously rated the perceived emotion of works each modified to express happy, sad, angry, tender, and normal. Intended emotion was identified correctly at 78%, with valence and arousal significantly shifted regardless of the works' original emotions. Existing systems developed for expressive performance, such as Director Musices (DM), focus on modifying features of performance. To study emotion more broadly, CMERS modifies features of both score and performance.

In Experiment 2, 18 participants rated music works modified by CMERS and DM to express five emotions. CMERS's intended emotion was correctly identified at 71%, DM at 49%. CMERS achieved significant shifts in valence and arousal, DM in arousal only. These results suggest that features of the score are important for controlling valence. The effects of musical training on emotional identification accuracy are also discussed.

Computer Music Journal, 34:1, pp. 41–64, Spring 2010

Background

> [E]verything in the nature of musical emotion that the musician conveys to the listener can be recorded, measured, repeated, and controlled for experimental purposes; and . . . thus we have at hand an approach which is extraordinarily promising for the scientific study of the expression of musical emotion (Seashore 1923, p. 325).

Empirical studies of emotion in music constitute one of the most practical resources for the development of a rule-based system for controlling musical emotions. For over a century, music researchers have examined the correlations between specific musical features and emotions (Gabrielsson 2003). One well-known example in the Western tradition is the modes' strong association with valence: major mode is associated with happy, and minor mode is associated with sad (Hevner 1935; Kastner and Crowder 1990). Although many exceptions to this rule exist in Western music literature, such a connection may have a cross-cultural basis. Recently, Fritz et al. (2009) showed that members of a remote African ethnic group who had never been exposed to Western music exhibited this association.

In the 1990s, the capability of musical features to be manipulated in the expression of different basic emotions received considerable interest. In a study of composition, Thompson and Robitaille (1992) asked musicians to compose short melodies that conveyed six emotions: joy, sorrow, excitement, dullness, anger, and peace. The music was performed in a relatively deadpan fashion by a computer sequencer. Results found that all emotions except anger were accurately conveyed to listeners. In a similar study of performance, Gabrielsson (1994, 1995) asked performers to play several well-known tunes, each with six different emotional intentions. Performers were found to vary the works' overall tempo, dynamics, articulation, and vibrato in relation to the emotion being expressed. Subsequent studies of performance found that both musicians and non-musicians could correctly identify the set of basic emotions being expressed (Juslin 1997a, 1997b).

Music may use an emotional "code" for communication. In this model, emotions are first encoded by composers in the notated score using the variation of musical features. These notations are then interpreted and re-encoded by performers in the acoustic signal using similar variations. These intentions are then decoded by listeners as a weighted sum of the two (Kendall and Carterette 1990; Juslin and Laukka 2004; Livingstone and Thompson 2009). This code is common to performers and listeners, with similar acoustic features used when encoding and decoding emotional intentions (Juslin 1997c). The code appears to function in a manner similar to that observed in speech and facial expression (Ekman 1973; Scherer 1986). Most recently, facial expressions in emotional singing may also use this code. Livingstone, Thompson, and Russo (2009) reported that singers' emotional intentions could be identified from specific facial features. Speech and music share many of the same features when expressing similar emotions (Juslin and Laukka 2003). The systematic modification of these features can bring about similar shifts in emotion (Peretz, Gagnon, and Bouchard 1998; Ilie and Thompson 2006). Livingstone and Thompson (2006, 2009) proposed that music may function as just one instantiation of a shared audio-visual emotional code that underlies speech, facial expression, dance, and the broader arts.

Leveraging this emotional code is a central component of computational models for controlling musical emotion. Two previous systems have been developed that use this approach: KTH's Director Musices (Friberg 1991; Bresin and Friberg 2000a; Friberg, Bresin, and Sundberg 2006), and CaRo (Canazza et al. 2004). Both of these rule-based systems were originally designed for automated expressive performance. This area of research is concerned with the generation of a natural ("humanistic") performance from a notated score. Many systems address this problem: GROOVE (Mathews and Moore 1970), POCO (Honing 1990), Melodia (Bresin 1993), RUBATO (Mazzola and Zahorka 1994), Super Conductor (Clynes 1998), SaxEx (Arcos et al. 1998), and the system by Ramirez and Hazan (2005). Here, *expressiveness* refers to the systematic deviations by the performer "in relation to a literal interpretation of the score" (Gabrielsson 1999, p. 522).

Both DM and CaRo have focused on the control of performance features when changing emotion. These systems do not modify particular aspects of the score, such as pitch height and mode. This choice reflects the different goals of the three systems, where DM and CaRo excel in their respective domains. However, in the investigation of emotion, such features play a central role. Although a recent extension to DM has proposed changes to the score, its effectiveness has not been examined (Winter 2006). As the score plays a central role in Western music in determining musical emotion (Thompson and Robitaille 1992; Gabrielsson and Lindstrom 2001), this focus on performance limits the utility of these two systems for research on emotion. More recently, Oliveira and Cardoso (2009) began investigating modifications to the score.

The capacity to isolate and control specific features provides researchers with a powerful methodology for exploring the perceptual relationships between music and emotion (Seashore 1923; Juslin and Laukka 2004). What types of functions govern these relationships? Are they linear, sigmoid, monotonic, or higher-order polynomial? For example, is a change of 5 dB in a soft work equivalent to that in a loud work? Does a change in mode only affect valence? The power of this methodology is only beginning to be explored (Ilie and Thompson 2006).

Human performers are unsuitable for this task, however. The expressive conventions learned through daily performance become hard-coded, where advanced performers are often unable to produce expressionless "deadpan" performances when asked (Gabrielsson 1988; Palmer 1992). Consequently, isolated modification cannot be achieved, as related features would be unconsciously modified by these learned conventions. Manipulation with sound-editing software is also inadequate, where only a limited subset of features can be modified (e.g., pitch height, tempo, and loudness). To address this, we present CMERS, a computational rule system for the real-time control of perceived musical emotion that modifies features at both the score level and the performance level, while generating an expressive performance.

In this article, we first present a cumulative analysis of the rules underlying CMERS. We begin by collating 148 empirical studies of music and emotion. We then identify the features that are central to controlling musical emotion, before concluding with a list of features used to generate the expressive deviations that mimic human performance. After that, we describe the system architecture and implementation details of the rule set. Next, we present the results of two perceptual experiments. In Experiment 1, we assess the capability of CMERS to change the perceived emotion of music. In Experiment 2, we compare CMERS with DM to highlight the importance of particular score features in changing musical emotion. We then discuss the outcomes of system testing and implications for the field, and we conclude with an outlook for future work.

Music-Emotion Rules and Expressive Performance Features

The relationships between musical features and emotion are typically referred to as "cues" (Brunswik 1952, 1956). However, this term lacks the necessary specificity in music-emotion research. We use the term *music-emotion rule* to represent the application and variation of a musical feature to bring about a specific change in musical emotion. A music-emotion rule has a *type*, a *variation*, and a set of *emotional consequents*. A set of emotional consequents is required as a single feature is often reported in the literature for more than one emotion. An example of a music emotion rule is: "Mode minor $\approx$ sad, angry." In this example, the rule *type* is "Mode," the *variation* is "minor," and the *emotional consequents* are "sad" or "angry." Thus, the minor mode is often used to convey "sadness" or "anger." A single rule type has multiple variations, e.g., Mode major or Mode minor. Each is a separate rule with a distinct set of emotional consequents.

Research into music-emotion rules has been divided largely into two camps, investigating both structure (the score; Gabrielsson and Lindstrom 2001) and performance (Gabrielsson 1999, 2003; Juslin and Laukka 2003). Most of these studies focused specifically on Western classical music. Features in the score that relate to emotion are the "factors in the composed musical structure represented in the musical notation" (Gabrielsson and Lindstrom 2001). We use the term *structural music-emotion rules* to refer to the subset of features notated in a score by the composer for the expression of particular emotions. Features in performance are those elements under the control of the performer that are conveyed, modified, or added through their interpretation of the notated score (Palmer 1997; Juslin 2001). We use the term *performance music-emotion rules* to refer to the subset of modifications or additions to the notated score by the performer for the expression of particular emotions. However, many performance features do not correlate with any specific emotion, and instead they may operate to accent the underlying structure. Systems for automated expressive performance use this feature set to mimic a human performance. We refer to these features as *expressive performance features*.

Some features can be modified by composers and performers and thus appear in both structural and performance music-emotion rule sets. Set membership is relaxed in some instances, as no unifying set of definitions has yet been proposed. Broadly, the taxonomy employed here is used to highlight the distinct stages in the communication process (Livingstone and Thompson 2009).

Rule Collation

Several cumulative analyses of music features and emotion have appeared in recent years; together, these provide a cohesive insight into the formidable number of empirical studies. Four such cumulative analyses are examined in this article. In each, the authors provided a single representation of emotion to categorize the results. A single representation was required, as the individual studies reviewed each employed their own emotional representation or terminology. This resulted in dozens of highly related emotional descriptions for a single feature. (For a review, see Schubert 1999a; Gabrielsson and Lindstrom 2001.)

A variety of emotion representations exist: open-ended paragraph response, emotion checklists, rank-and-match, rating scales, and dimensional representations (see Schubert 1999a). In this article, we have adopted a two-dimensional circumplex model of emotion (Russell 1980), specifically Schubert's (1999b) Two-Dimensional Emotion Space (2DES). This representation does not limit emotions and user responses to individual categories (as checklist and rank-and-match do), but it does allow related emotions to be grouped into four quadrants (see Figure 1). The quantitative nature of the 2DES lends itself to a computational system, unlike the paragraph response, checklist, and rank-and-match measures, by permitting a fine-grained coupling between rule variation parameters and emotions. Finally, the 2DES allows for a continuous response methodology in user testing (Schubert 2001).

Although no representation can capture the breadth and nuanced behavior of human emotion, the goal is to strike a balance between a richness of description and the limitations imposed by empirical study. A discussion of the selection process, translations between the representations, and all rule definitions is found in Livingstone (2008).

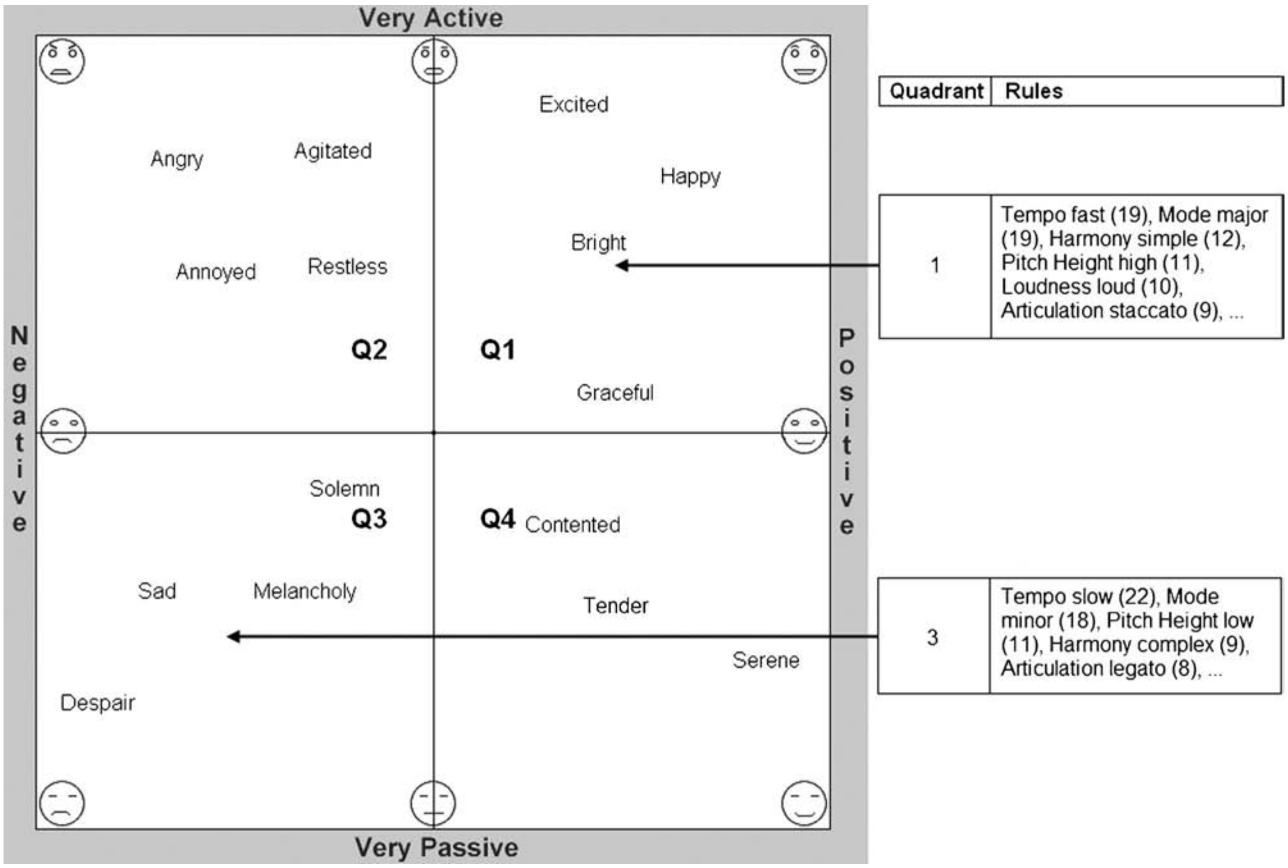
Structural and Performance Music-Emotion Rules

Table 1 provides a cumulative analysis of 102 unique studies of structural music-emotion rules (Schubert 1999a; Gabrielsson and Lindstrom 2001). The table is grouped by the quadrants of the 2DES, which are referred to loosely as (1) happy, (2) angry, (3) sad, and (4) tender. An example of how these rules map onto the 2DES is illustrated in Figure 1.

The rule variations in Table 1 have inter-agreement of 88%, with only 47 of the 387 cited rule variations conflicting. This score is a ratio of conflicting citations (numerator) to agreeing citations (denominator) in a single quadrant, added over all quadrants. For example, in Quadrant 1: 19 studies cited Tempo fast, 1 study cited Tempo slow, 12 studies cited Harmony simple, and 2 cited Harmony complex $= (1 + 2 + \cdots)/(19 + 12 + \cdots)$. This low rate of conflict between rule variations indicates a high degree of cross-study agreement. Six rule types (as opposed to rule variations) were identified as appearing in all quadrants with three or more citations: Tempo (65), Mode (62), Harmonic Complexity (47), Loudness (38), Pitch Height (31), and Articulation (25). Their occurrence in all quadrants suggests they are particularly useful for controlling emotion. Other common rule types with three or more citations included Pitch Range (11) and Pitch Contour (11).

Similarly, Table 2 provides a cumulative analysis of 46 unique studies of performance (as opposed to structural) music-emotion rules (Schubert 1999a; Juslin 2001; Juslin and Laukka 2003). An additional performance rule type not listed in Table 2 is

Figure 1. Mapping of structural music-emotion rules for Quadrants 1 and 3 of the 2DES. For example, Tempo fast maps to Quadrant 1 (emotional consequent).



Expressive Contours (Juslin 2001). This rule is discussed later in more detail.

Table 2 rule variations have inter-agreement of 87% (53 conflicts out of 401 total citations), again indicating a high degree of cross-study agreement. Five rule types (as opposed to rule variations) were identified as occurring in all quadrants with three or more citations: Tempo (61), Loudness (54), Articulation (43), Note Onset (29), and Timbre Brightness (27). Their occurrence in all quadrants suggests they are particularly useful for controlling emotion. Other common rule types with three or more citations included Articulation Variability (15) and Loudness Variability (14). In Table 2, we use the timbral terms *bright* and *dull* to encompass *sharp, bright,* and *high-frequency energy,* and *soft, dull,* and *low-frequency energy*, respectively. Recently, Mion et al. (2010, this issue) identified Roughness and Spectral Centroid as useful performance features for controlling perceived emotion. This extends previous work on these features (Schubert 2004).

Primary Music-Emotion Rules

Particular rule types in Tables 1 and 2 appeared in all quadrants of the 2DES, but with differing rule variations. To understand this behavior, these rule types and their set of variations were mapped onto the 2DES (see Figure 2). Rule instances near the periphery have the highest citation count. For example, "happy" music in Quadrant 1 typically has a faster tempo (#1, 19 structural and 18 performance studies), a major mode (#2, 19 structural), simple harmonies (#3, 12 structural), higher loudness (#4, 10 structural and 8 performance), staccato articulation (#5, 9 structural and 12 performance), above-average

Table 1. Structural Music-Emotion Rules

Quadrant/Emotion	*Structural Music-Emotion Rules*
1 Happy	**Tempo fast (19), Mode major (19), Harmony simple (12), Pitch Height high (11), Loudness loud (10), Articulation staccato (9), Pitch Range wide (3), Pitch Contour up (3)**, Pitch Height low (2), Pitch Variation large (2), Harmony complex (2), Rhythm regular (2), Rhythm irregular (2), Rhythm varied (2), Rhythm flowing (2), Loudness Variation small (1), Loudness Variation rapid (1), Loudness Variation few (1), Note Onset rapid (1), Pitch Contour down (1), Timbre few (1), Timbre many (1), Tempo slow (1), Rhythm complex (1), Rhythm firm (1), Tonality tonal (1)
2 Angry	**Harmony complex (16), Tempo fast (13), Mode minor (13), Loudness loud (10), Pitch Height high (4), Pitch Height low (4), Articulation staccato (4), Articulation legato (4), Pitch Range wide (3)**, Pitch Contour up (2), Pitch Variation large (2), Loudness Variation rapid (2), Rhythm complex (2), Note Onset Rapid (1), Note Onset slow (1), Harmony simple (1), Rhythm irregular (1), Pitch Contour down (1), Timbre many (1), Tempo slow (1), Rhythm firm (1), Loudness soft (1), Loudness Variation large (1), Pitch Variation small (1), Timbre sharp (1), Tonality atonal (1), Tonality chromatic (1)
3 Sad	**Tempo slow (22), Mode minor (18), Pitch Height low (11), Harmony complex (9), Articulation legato (8), Loudness soft (7), Pitch Contour down (4), Harmony simple (3), Pitch Range narrow (3)**, Note Onset slow (2), Pitch Variation small (2), Rhythm firm (2), Mode major (2), Tempo fast (1), Loudness loud (1), Loudness Variation rapid (1), Loudness Variation few (1), Pitch Height high (1), Pitch Contour up (1), Rhythm regular (1), Timbre many (1), Tonality chromatic (1), Timbre few (1), Timbre soft (1)
4 Tender	**Mode major (12), Tempo slow (11), Loudness soft (11), Harmony simple (10), Pitch Height high (5), Pitch Height low (4), Articulation legato (4), Articulation staccato (4), Tempo fast (3)**, Pitch Contour down (2), Pitch Range narrow (2), Mode minor (1), Pitch Variation small (1), Note Onset slow (1), Pitch Contour up (1), Loudness Variation rapid (1), Rhythm regular (1), Loudness Variation few (1), Timbre few (1), Timbre soft (1), Rhythm flowing (1), Tonality tonal (1)

Rules are grouped by emotional consequent and list all rule "Type variation" instances for that emotion. Numbers in parentheses indicate the number of independent studies that reported the correlation. Those in **bold** were reported by three or more independent studies. Adapted from Livingstone and Brown (2005).

pitch height (#6, 11 structural), fast note onsets (#7, 5 performance), and moderate to high timbral brightness (#8, 8 structural).

Figure 2 displays a striking relationship between rule variations and their emotional consequents. All rule variations alternate between the expression of high versus low arousal, or positive versus negative valence. For example, to express high arousal in music, the tempo should be increased, loudness increased, the articulation should be made more staccato, pitch height raised, and timbral brightness should be increased. Conversely, to express low arousal, the tempo should be decreased, loudness decreased, articulation made more legato, pitch height lowered, and timbral brightness decreased. This reflective symmetry suggests these musical features function as a code for emotional communication in music (Livingstone and Thompson 2009).

This set has been termed the *Primary Music-Emotion Rules*, and it consists of rule types Tempo, Mode, Harmonic Complexity, Loudness, Articulation, Pitch Height, Note Onset, and Timbre Brightness. We consider these eight rule types fundamental to the communication of emotion in Western classical music. Only two rule types in Figure 2 have been identified as controlling the valence of a work: Mode and Harmonic Complexity. This suggests that control of one or both of these score features is crucial to changing the valence of a work. This hypothesis is revisited in Experiment 2.

Table 2. Performance Music-Emotion Rules

Quadrant/Category	*Performance Music-Emotion Rules*
1 Happy	**Tempo fast (18), Articulation staccato (12), Loudness medium (10), Timbre medium bright (8), Articulation Variability large (6), Loudness loud (5), Note Onset fast (5), Timing Variation small (5), Loudness Variability low (4), Pitch Contour up (4), Microstructural Regularity regular (4), F0 sharp (3)**, Vibrato fast (2), Vibrato large (2), Pitch Variation large (2), Loudness Variability high (2), Duration Contrasts sharp (2), Duration Contrasts soft (2), Tempo medium (1), Articulation legato (1), Note Onset slow (1), Vibrato small (1), Pitch Variation small (1), Loudness low (1), Loudness Variability medium (1), Timbre bright (1), Timing Variation medium (1)
2 Angry	**Loudness loud (18), Tempo fast (17), Articulation staccato (12), Note Onset fast (10), Loudness low (9), Timbre bright (8), Vibrato large (7), F0 sharp (6), Loudness Variability high (6), Timbre dull (5), Microstructural Regularity irregular (5), Tempo medium (4), Articulation legato (4), Articulation Variability large (4), Articulation Variability medium (4), Duration Contrasts sharp (3)**, Vibrato fast (2), Vibrato small (2), Loudness variability low (2), Microstructural Regularity regular (2), Timing Variation medium (2), Timing Variation small (2), Tempo slow (1), Pitch Variation large (1), Pitch Variation medium (1), Pitch Variation small (1), F0 precise (1), F0 flat (1), Loudness medium (1), Pitch Contour up (1)
3 Sad	**Tempo slow (18), Loudness low (16), Articulation legato (12), F0 flat (11), Note Onset slow (9), Timbre dull (7), Articulation Variability small (5), Vibrato slow (3), Vibrato small (3), Timing Variation medium (3), Pitch Variation small (3), Duration Contrasts soft (3)**, Loudness Variability low (2), Microstructural Regularity irregular (2), Timing Variation large (2), Tempo medium (1), Vibrato large (1), Loudness medium (1), Loudness Variability high (1), Pitch Contour down (1), Microstructural Regularity regular (1), Timing Variation small (1)
4 Tender	**Loudness low (10), Tempo slow (8), Articulation legato (7), Note Onset slow (5), Timbre dull (4), Microstructural Regularity regular (3), Duration Contrasts soft (3)**, Loudness Variability low (2), Timing Variation large (2), Vibrato slow (1), Vibrato small (1), Pitch Variation small (1), Pitch Contour down (1)

Rules are grouped by emotional consequent and list all rule "Type variation" instances for that emotion. Numbers in parentheses indicate the number of independent studies that reported the correlation.

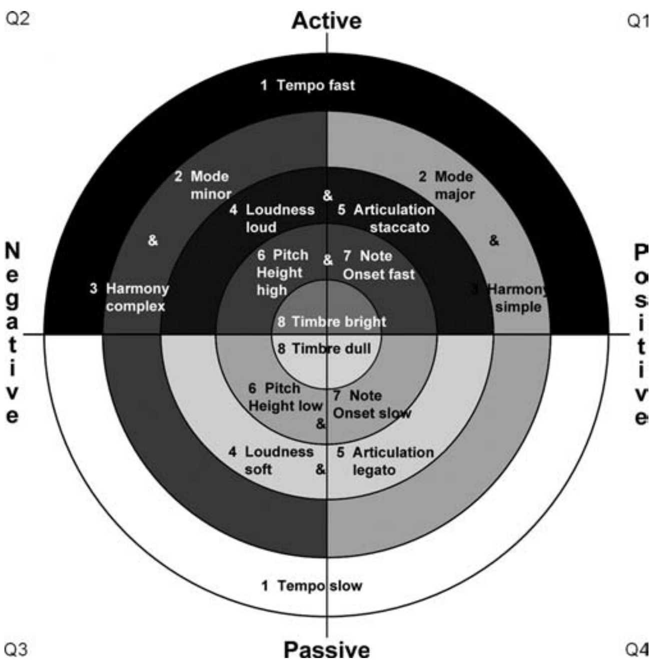
Expressive Performance Features

Expressive performance features are not used to change musical emotion. However, they are needed for any system attempting this task. Earlier pilot testing by Livingstone and Brown (2005) found that whereas emotion could be changed without these features, participants reported difficulty with the concept, as the music was performed "mechanically." Three cumulative studies were used in the collation of expressive performance features (Gabrielsson 1999, 2003; Lindström 2004). Eleven expressive performance features were selected and were grouped into six *expressive feature rules*, listed in Table 3. The Expressive Phrase Curve is typically treated as an expressive performance feature; however, in particular circumstances, it can also function as a music-emotion rule. Bresin and Friberg (2000b) have noted that an inverted arch is effective at communicating anger. This logic is also applied in CMERS.

CMERS System Architecture

CMERS is designed as a filter-based system for real-time MIDI modification, as shown in Figure 3. The system is implemented in the programming language Scheme, and it uses the Impromptu music-programming environment (Sorensen and Brown 2007). On start-up, MIDI data and markup files are entered into CMERS, which converts them into native Scheme data structures. Expressive

Figure 2. The set of Primary Music-Emotion Rules mapped on to the 2DES. Adapted from Livingstone and Thompson (2006).



feature rules that can be applied pre-runtime are executed.

At runtime, MIDI data is passed through a series of real-time filters. Each filter is responsible for a specific music-emotion rule type. When the filter is engaged, it applies the specified rule variation; when disengaged, it makes no modification. For example, a Mode filter that is engaged modifies notes to be in either the major mode or the minor mode, while a Loudness filter modifies the intensity of the note ($\pm n$ dB). Individual filters are aggregated into four *filter control sets* that correspond to the quadrants of the 2DES. A 2DES interface is used by the system operator to select the desired emotion of the musical output. This action invokes the corresponding filter control set, which engages all member filters and passes them the rule-variation parameters corresponding to where in the quadrant was selected. These filters in turn modify note information in real-time, prior to being scheduled for playing. CMERS possesses a variable-length scheduling buffer to allow for smooth operation on older machines. CMERS was the first system to possess real-time music emotion-modification capability (Livingstone and Brown 2005), which DM now also supports via the pDM extension (Friberg 2006). CMERS uses the Akai Steinway 3 SoundFont for real instrument sound.

The abstraction provided by filter control sets allows for the modification of music to occur at the level of emotion. This design philosophy enables the integration of CMERS into a variety of applications, such as computer gaming, adaptive music environments, and electronic music mixing (Livingstone, Brown, and Muhlberger 2005; Livingstone et al. 2007).

Additional Markup

CMERS requires additional music markup that is provided by the system operator. First, *Slur Membership* is a ternary value indicating a note's position in the slur (onset, within, offset), needed for the Slur expressive feature rule. Slur information is taken from the notated score. As MIDI does not code slur information, it must be added. Second, *Melody/Harmony Membership* is a binary value indicating if the note belongs in the melody or harmony voice; it is needed for Global Melody Accent and Pedal expressive feature rules. Third, *Note Signature* is the key in which the current note lies. As composers frequently modulate away from the work's notated key, this data is required for correct operation of the Mode music-emotion rule. Fourth, *Phrase Boundaries* specify the hierarchical phrase structure of the music. This information is used in the Expressive Phrase Curve feature rule (discussed next).

The goal of CMERS is to provide researchers with a tool for testing the relationships between musical features and emotion. While CMERS is capable of generating an expressive performance from a deadpan score, this is not the primary goal of the system. The inclusion of markup was justified to meet project time constraints. Future releases may incorporate automatic determination algorithms (for example, Krumhansl and Kessler 1982).

Music Object Hierarchy

A music work is represented in CMERS using the *music object hierarchy*, illustrated in Figure 4. The

Table 3. Expressive Performance Features Implemented in CMERS

Expressive Feature Rules	*Expressive Performance Feature*	*Description*
Expressive Phrase Curve	Rubato, Ritardando, and Final Ritard	Rubato and ritardando are a fundamental aspect of performance, and involve the expressive variation of tempo (Gabrielsson 2003).
	Loudness Strategies	Each performer employs a unique, weighted combination of loudness profiles (Gabrielsson 2003). Profiles are highly consistent across performances.
	Accent and Loudness	Performer accents typically highlight phrase hierarchy boundaries rather than higher intensity (climax) passages (Gabrielsson 1999).
Pedal	Pedal and Legato Articulation	Pianists employ the sustain pedal during a typical classical performance, with higher usage during legato sections (Gabrielsson 1999).
Chord Asynchrony	Chord Asynchrony	Chord notes are played with temporal onset-offset asynchrony. Degree of asynchrony is coupled with phrase hierarchy boundaries, indicating its use as an accent device (Gabrielsson 1999).
	Accented Loudness	Increased loudness for melody lead note (Goebl 2001). Intensity is coupled with phrase hierarchy boundaries (Livingstone 2008).
Slur	Slur	Notes are grouped into a coherent unit to convey low-level phrase structure (Sundberg, Friberg, and Bresin 2003; Chew 2007).
Metric Accent	Metric Accent	Loudness and timing accents are used to communicate the metrical structure of the work (Gabrielsson 1999; Lindström 2004).
Global Melody Accent	Global Melody Accent	In homophonic music, the melody is typically more salient than the accompaniment, and is usually played with greater intensity (Repp 1999).
(Entailed by multiple rules)	Exaggerated Performance	Exaggerated performance results in increased chord asynchrony, rubato, and articulation (Gabrielsson 1999). Has not been correlated with any particular emotion.
(Deterministic system)	Performer Consistency	High within-individual consistency across performances of the same work.

hierarchy is based on GTTM's Grouping Structure (Lerdahl and Jackendoff 1983) and is automatically generated from the phrase boundary markup and MIDI file. The hierarchy links phrases, chords, and notes into a single structure and simplifies the application of music-emotion rules and expressive-feature rules. The Expressive Phrase Curve generates hierarchy weights for individual phrases pre-runtime that can be rapidly accessed for specific rule filters at runtime (see Table 4).

Music-Emotion Rules Implemented in CMERS

Eight music-emotion rules types were selected for implementation and are listed in Table 4. The selection of rules was based on the Primary

Figure 3. High-level architecture of CMERS.

Figure 4. Music object hierarchy used in CMERS. An example hierarchy for a music sample consisting of the first twelve bars of Mozart's Piano Sonata No. 12. A larger excerpt would involve additional hierarchical levels.

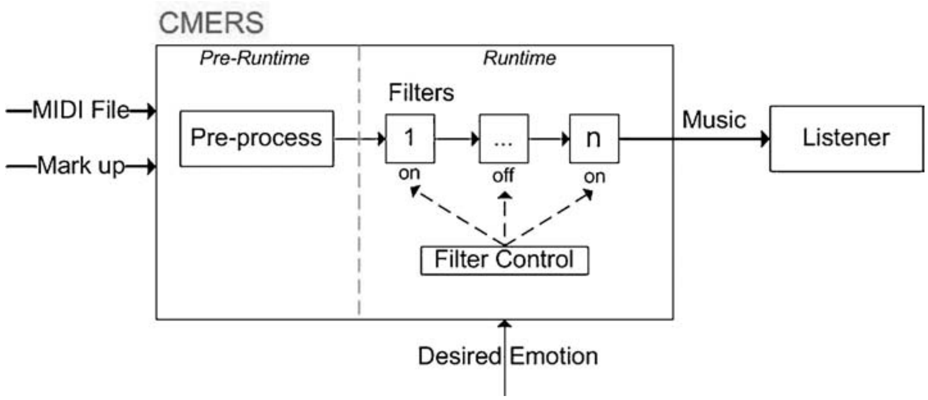


Figure 3.

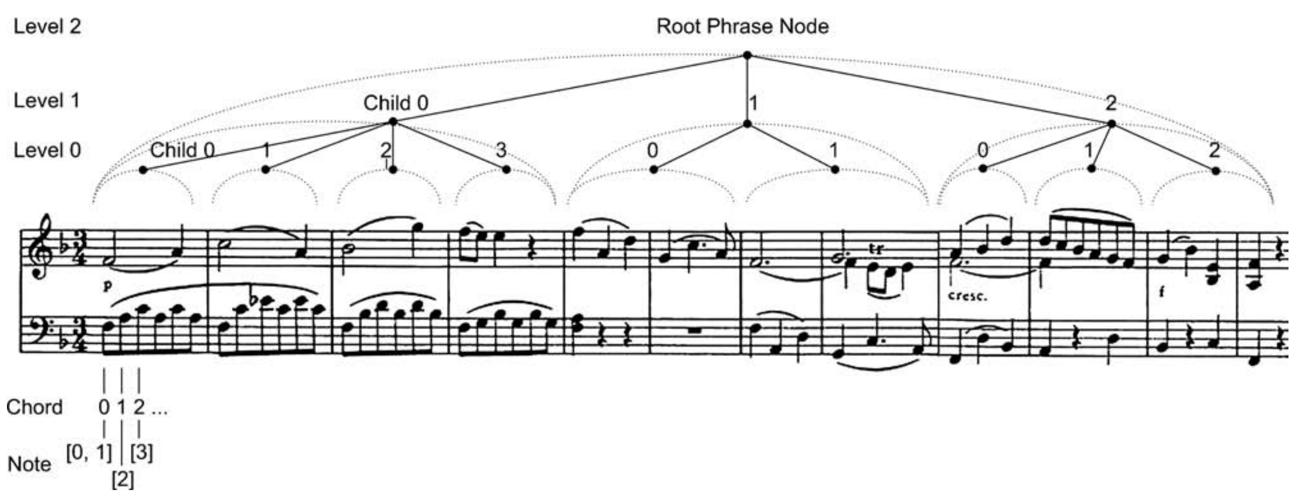


Figure 4.

Music-Emotion Rules (see Figure 2), and additional project constraints. In Table 4, "average phrase weight" refers to the average Level 0 Expressive Phrase Curve weight across all chords in that phrase. This enables articulation and loudness to vary with the specified Level 0 phrase intensity (see the subsequent Expressive Phrase Curve discussion). Additionally, notes occurring within a slur have their articulation variability reduced by 30% for happy and 50% for anger.

Although CMERS supports polyphony and multiple concurrent instruments, this implementation focused on solo piano. Selected rules and features reflect this choice (i.e., no timbral change). Rule-variation parameters (e.g., *n* dB louder, or *n* BPM slower) were generated using analysis-by-synthesis (Sundberg, Askenfelt, and Frydén 1983; Gabrielsson 1985). This methodology details a nine-step iterative process for locating features for study (analysis), synthesizing values for these rules, and then having listeners judge the musical quality of these values. Judgments of quality for CMERS were performed by the first three authors and two other musicians. All individuals had extensive musical training; three also had extensive music teaching experience. Two of the Primary Music-Emotion Rule types—Harmonic Complexity and Note Onset—were not implemented in this version of CMERS owing to

Table 4. Music-Emotion Rules Implemented in CMERS

Quadrant/Emotion	*Rule Type*	*Variations*	*Details*
1 Happy	Tempo	Increase	10 BPM
	Mode	Major	
	Loudness	Increase	5 dB
	Pitch Height	Raise	4 semitones
	Expressive Phrase Curve	Normal	
	Expressive Contour ::		
	Articulation	Staccato	Average ≈ 75% (+average phrase weight)
	Articulation variability	Large	96% → 55% (+average phrase weight)
	Loudness variability	Moderate	96% → 70% (+average phrase weight)
2 Angry	Tempo	Increase	10 BPM
	Mode	Minor	
	Loudness	Increase	7 dB
	Pitch Height	Unchanged	
	Expressive Phrase Curve	Inverse	
	Expressive Contour ::		
	Articulation	Staccato	Average ≈ 80% (+average phrase weight)
	Articulation variability	Large	85% → 93% → 55% (+ avg. phrase weight)
	Loudness variability	Large	85% → 93% → 55% (+ avg. phrase weight)
3 Sad	Tempo	Decrease	15 BPM
	Mode	Minor	
	Loudness	Decrease	5 dB
	Pitch Height	Lower	4 semitones
	Expressive Phrase Curve	Normal	
	Expressive Contour ::		
	Articulation	Legato	Average ≈ 93% (+average phrase weight)
	Articulation variability	Low	90% → 98% → 85% (+ avg. phrase weight)
	Loudness variability	Low	85% → 93% → 55% (+ avg. phrase weight)
4 Tender	Tempo	Decrease	20 BPM
	Mode	Major	
	Loudness	Decrease	7 dB
	Pitch Height	Raise	4 semitones
	Expressive Phrase Curve	Normal	
	Expressive Contour ::		
	Articulation	Legato	Average ≈ 90% (+average phrase weight)
	Articulation variability	Low	95% → 85% (+ average phrase weight)
	Loudness variability	Low	95% → 85% (+ average phrase weight)

constraints and technical problems. Two rule types require additional explanation.

Mode Rule Type

The Mode rule type is unique to CMERS and is not present in existing rule systems for changing musical emotion (DM and CaRo). The mode rule converts a note into those of the parallel mode. No change in Pitch Height occurs when converting to the parallel mode, as this is a separate rule type. In converting major to minor, the third and sixth degrees of the diatonic scale are lowered a semitone, and conversely, they are raised in the minor-to-major conversion. For simplicity, the seventh degree was not modified. This version of the Mode rule type only functions for music in the Harmonic major or Harmonic minor diatonic scale. Future versions could be extended to handle ascending and descending melodic minor scales.

Expressive Contour Rule Type

The Expressive Contours Rule type applies varying patterns of articulation to Level 0 phrases (see Figure 4), and it subsumes the role of a global articulation rule. It was hypothesized that global articulation values commonly reported in the performance literature (e.g., staccato and legato) obfuscated the underlying articulation patterns of performers. Based on work by Juslin (2001), Quadrants 1 and 4 possess a linear incremental decrease in note duration over a Level 0 phrase, and Quadrants 2 and 3 possess an inverted "V" shape over a Level 0 phrase. Quadrants 1 and 2 possess steeper gradients than Quadrants 3 and 4. Note articulation is the ratio of the duration between the duration from the onset of a tone until its offset and the duration from the onset of a tone until the onset of the next tone (Bengtsson and Gabrielsson 1980). Table 4 lists duration ratios as percentages under Articulation and Articulation Variability.

Expressive Features Rules Implemented in CMERS

All six expressive feature rules listed in Table 3 were implemented in CMERS. We now describe the implementation detail of those feature rules.

Expressive Phrase Curve

The Expressive Phrase Curve is modeled on Todd's theory of expressive timing (1985, 1989a, 1989b) and dynamics (1992) in which a performer slows down at points of stability to increase the saliency of phrase boundaries. This behavior underlies DM's Phrase Arch rule (Friberg 1995).

The curve represents the chord/note's importance within the phrase hierarchy and is used in Expressive Phrase Curve, Expressive Contour, and Chord Asynchrony, which together affect six expressive performance features. CMERS generates Tempo and Loudness modification curves using a second-order polynomial equation (Kronman and Sundberg 1987; Repp 1992; Todd 1992; Friberg and Sundberg 1999; Widmer and Tobudic 2003). Individual curves are generated for every phrase contained within the markup, at each level in the hierarchy. For example, the curves generated for Figure 4 consist of nine Level 0 curves, three Level 1 curves, and one Level 2 curve. These curves are added vertically across all hierarchal levels to produce a single Expressive Phrase Curve. Therefore, a chord's tempo and loudness modifier is the sum of all phrase curves (at that time point) of which it is a member in the hierarchy. In Figure 4, for example, Chord 0's modifier is the sum of curve values at time $t = 0$ for Levels 0, 1, and 2.

The three coefficients of a second-order polynomial $(ax^2 + bx + c)$ can be solved for the desired inverted arch with $a < 0$, $b = 0$, $c > 0$. A single phrase corresponds to the width of the curve, which is the distance between the two roots of the polynomial: $x = w/2$, where w is the width. The polynomial is then solved for a when $b = 0 : (-b \pm (b^2 - 4ac)^{1/2})/2a$. Solving for a yields $a = -4c/w^2$. As c represents the turning point of the curve above the x-axis, $c = h$, where h is the height. As we need a curve that slows down at the roots and speeds up at the turning point, the curve is shifted halfway below the x axis by subtracting $h/2$. Substituting into $y = ax^2 + c$ yields Equation 1:

$$y_i = \sum_{n=0}^{N} \frac{-4h_n x_n^2}{w_n^2} + \frac{h_n}{2} \tag{1}$$

where y_i is the tempo and loudness-curve modifier of chord i, x is the order of the chord in phrase n (e.g., 0, 1, or 2 in Figure 4), h is the maximum intensity of that curve (variable for each hierarchy level), w is the width of phrase n, and N is the number hierarchy levels starting from 0 (e.g., $N = 2$ in Figure 4). To generate a work's Expressive Phrase Curve, a single phrase height h is required for each level of the hierarchy. For example, Figure 4 would require three values (Levels 0, 1, and 2). As the music-object hierarchy is a tree structure, CMERS employs a recursive pre-order traversal algorithm to apply the expressive-phrase curve modifier to each note.

Analyses of performer timing typically report greater lengthening at phrase endings (Palmer 1989; Todd 1985; Repp 1992; see also Todd 1992). The Expressive Phrase Curve, however, implements a symmetrical form of rubato. This simplified model was chosen to reduce required markup. Analysis-by-synthesis reported pleasing results with this model.

Pedal

A pedal-on event is called for the first note of a slur occurring in the melody line, and pedal-off is called for the last note. This model does not capture the full complexity of pedaling in Western classical performance, but it was adopted for reasons of parsimony. However, analysis-by-synthesis reported pleasing musical results for the experiment stimuli, with only one "unmusical" event.

Chord Asynchrony

Melody lag and loudness accent is the default setting, with a variable onset delay of of 10–40 msec and a variable loudness increase of 10–25%. The delay and loudness increase are determined by the chord's importance in the phrase hierarchy (see Equation 1). Optimal results were obtained when asynchrony was applied for the first chord in phrase Level 1. Following Goebl (2001), the next version of CMERS will default to having the melody note lead the other chord tones.

Slur

The slur is modeled to mimic the "down-up" motion of a pianist's hand. This produces increased loudness (110%) and duration (100%) for the first note, increased duration of all notes within the slur ($100 \pm 3\%$), and decreased loudness (85%) and duration ($70 \pm 3\%$) for the final note (Chew 2007).

Metric Accent

Metric accent employs a simplified form of GTTM's metrical structure (Lerdahl and Jackendoff 1983), by increasing the loudness of specified notes in each bar based on standard metric behaviors. For example, in 4/4 time, the beats are strong (+10%), weak, medium (+5%), weak. Although more sophisticated methods of metrical analysis exist (Temperley 2004), this behavior was sufficient for the current system.

Global Melody Accent

The global melody accent rule dictates a decrease in the loudness of all harmony notes by 20%.

Discussion

The system that we have outlined provides a parsimonious set of rules for controlling performance expression. Our rules undoubtedly simplify some of the expressive actions of musicians, but they collectively generate a believably "human" performance. Informal feedback from Experiment 1 found that listeners were unaware that the music samples were generated by a computer. We tentatively concluded that these decisions were justified for the current project.

Rule Comparison of CMERS and Director Musices

CMERS and DM share similar music-emotion rules. This result is unsurprising, as both utilize the findings of empirical music emotion studies. From Table 4, CMERS modifies six music features: mode, pitch height, tempo, loudness, articulation, and timing deviations. Timing deviations refers to rubato and final ritardando, achieved with the Expressive Phrase Curve rule. DM modifies four features: tempo, loudness, articulation, and timing deviations. Timing deviations refer to rubato, final ritardando, and tone group lengthening, achieved with the Phrase Arch and Punctuation rules (Bresin and Friberg 2000a).

Although the details of these rule sets differ, with DM possessing more sophisticated timing deviations, there is a qualitative difference between the two. As illustrated in Figure 2, only two rule types potentially control the valence of a musical work: mode and harmonic complexity. CMERS modifies the mode, DM does not modify either. This difference reflects the contrasting goals of the two systems. Subsequently, it is hypothesized that the two systems will differ markedly in their capacity to shift the valence of a musical work. This will be examined in Experiment 2. CMERS and DM also possess divergent expressive feature rules. However, as these features are not known to modify perceived emotion, they will not be examined here.

Experiment 1

The main hypothesis proposed for examination here is that CMERS can change the perceived emotion of all selected musical works to each of the 2DES quadrants (happy, angry, sad, and tender). The second and third hypotheses to be examined are that CMERS can successfully influence both the valence and arousal dimensions for all selected musical works.

Method

Participants

Twenty university students (12 women, 8 men), 18–23 years old (mean $\mu = 19.8$; standard deviation $\sigma = 1.54$) volunteered to participate and were each paid $25 AUD. Nine of the participants were enrolled as music students, seven possessed formal music training, and four were experienced listeners of Western classical music.

Musical Stimuli

Three stimuli were used: (1) Mozart's *Piano Sonata No. 12*, KV 332, in F Major (the first 40 bars of movement 1, unmodified, 48 sec duration); (2) Beethoven's *Piano Sonata No. 20*, Op. 49, No. 2 in G Major (the first 20 bars of movement 1, unmodified, 34 sec duration); and (3) Mendelssohn's *Songs without Words*, Op. 19 No. 2 in A Minor (the first 30 bars, unmodified, 44 sec duration). The three works provide a range of emotions to be modified, along with variation in period, structure, composer, key, and mode. In their unmodified states, the Beethoven work is characterized as happy, the Mendelssohn as sad, and the Mozart as beginning as happy before moving to angry/agitated at bar 22. Original MIDI files contained minimal dynamic and articulation information. Each file was processed by CMERS to obtain four emotionally modified versions and one emotionally unmodified version. Expressive feature rules were applied to all five versions to produce an expressive ("humanistic") performance. Output was recorded as 44.1 kHz, 16-bit, uncompressed sound files. All stimuli are available online at www.itee.uq.edu.au/~srl/CMERS.

Testing Apparatus

Continuous measurement of user self-reporting was used. A computer program similar to Schubert's (1999b) 2DES tool was developed (Livingstone 2008). This tool presents users with a dimensional representation of emotion on a 601-point scale of -100% to $+100\%$ (see Figure 1), and captures the user's mouse coordinates in real-time. Coordinate data was then averaged to produce a single arousal-valence value per music sample for each participant. Approximately $(20 \times 15 \times 420) = 126{,}000$ coordinate points were reduced to $(20 \times 15) = 300$ pairs.

Participant Handouts

An emotion handout was distributed which explained the difference between felt and perceived emotion (Gabrielsson 2002; Juslin and Laukka 2004; Livingstone, Brown, and Muhlberger 2005). Listeners were instructed to respond to what emotion they perceived the music to be expressing, rather than how the music made them feel. A second handout described arousal and valence, and how they combined to form the 2DES. Participants were told that the placement of emotion terms was a guide only, and that they should evaluate where on the axes they perceived the emotion to fall. A movie tutorial described the operation of the 2DES tool.

Procedure

Participants were tested individually with Sony MDR-P10 headphones, adjusted to a comfortable volume level. Four testing sessions were conducted. Participants heard 15 music samples, five versions of each music work (four that were emotionally modified along with the unmodified). Order was randomized across participants. Participants first listened to the sample, then opened the tool and

Figure 5. CMERS mean correctness for each emotion and musical work in Experiment 1.

heard the sample again while they continuously rated the emotion. This was done to familiarize participants, reducing the lag between music and rating onsets. Participants were allowed to repeat the sample if a mistake was made. Testing took approximately 50 minutes.

Data Analysis

For hypothesis 1 (that perceived emotion can be shifted to all 2DES quadrants), coordinate values were reduced to a binary measure of correctness—a rating of 1 if Perceived Emotion matched Intended Emotion, or 0 if not. A multinomial logistic regression was conducted to determine if a significant relationship existed between Perceived Emotion (four categories) and Intended Emotion (four categories). A binary logistic regression was conducted to determine if Correctness (the dependent variable) differed significantly over Music Work (an independent variable) and Intended Emotion (an independent variable). Given the use of binary data values, logistic regression is more appropriate than ANOVA.

For hypotheses 2 and 3 (that both valence and arousal can be shifted), two-way repeated-measures analyses were conducted with the fixed factors Intended Emotion (four levels) and Music (three levels), with coordinate points on the 2DES plane entered for arousal and valence values. Two contrast models were also produced with two additional two-way repeated measures, with fixed factors Intended Emotion (five levels: four quadrants, along with the unmodified version) and Music (three levels).

Results

The multinomial logistic regression analysis revealed that the interaction between Intended Emotion and Perceived Emotion was highly significant with $\chi^2(9) = 11183.0$, $p < .0005$, indicating that CMERS was successful in changing the perceived emotion of the music works. The binary logistic regression analysis revealed that the interaction between Music Work and Intended Emotion on Correctness was not significant with $\chi^2(6) = 11.91$, $p = .0641$. This indicates that the combination of Music Work and Intended Emotion did not affect Correctness. The effect of Music Work on Correctness was not significant with $\chi^2(2) = 1.59$, $p = .4526$, nor was the effect of Intended Emotion on Correctness with $\chi^2(3) = 1.51$, $p = .6803$. The emotion (Quadrant) accuracy means for each musical work are illustrated in Figure 5, with confusion matrices listed in

Table 5. Confusion Matrix of Emotion Identification (%) across the Three Works in Experiment 1

	Intended Quadrant/Emotion			
Perceived Emotion	*1/Happy*	*2/Angry*	*3/Sad*	*4/Tender*
1/Happy	**80**	15	1	15
2/Angry	13	**82**	12	3
3/Sad	2	3	**75**	7
4/Tender	5	0	12	**75**

Table 5. The mean accuracy across the four emotion categories for all three music works was 78%.

The two-way repeated measures for valence found that Intended Emotion was significant with $F(3, 57) = 75.561$, $p < .0005$, and for arousal Intended Emotion was significant with $F(3, 57) = 117.006$, $p < .0005$. These results indicate that CMERS was highly successful in influencing both valence and arousal dimensions. For valence, the interaction between Music Work and Intended Emotion was not significant, with $F(6, 14) = 0.888$, $p = .506$. For arousal, the interaction between Music Work and Intended Emotion was significant: $F(6, 114) = 5.136$, $p < .0005$. These results indicate that the choice of music affected arousal but not valence.

Plots of the change in valence and arousal intensities from the emotionally unmodified versions are illustrated in Figure 6. These graphs illustrate that significant modifications were made to both valence and arousal, with all music works pushed to similar intensities for each quadrant.

Two-way ANOVA analyses of arousal and valence with within-subjects contrast effects are listed in Table 6. These results illustrate that significant shifts in both valence and arousal away from the unmodified versions were achieved for all quadrants, across the three music works.

Experiment 2: CMERS and DM

The main hypothesis proposed for examination here is that CMERS is more accurate than DM is at changing the perceived emotion of selected musical works to all 2DES quadrants (happy, angry, sad, and tender). The second and third hypotheses to be examined are that CMERS is significantly more effective at influencing valence and arousal than DM.

Method

Participants

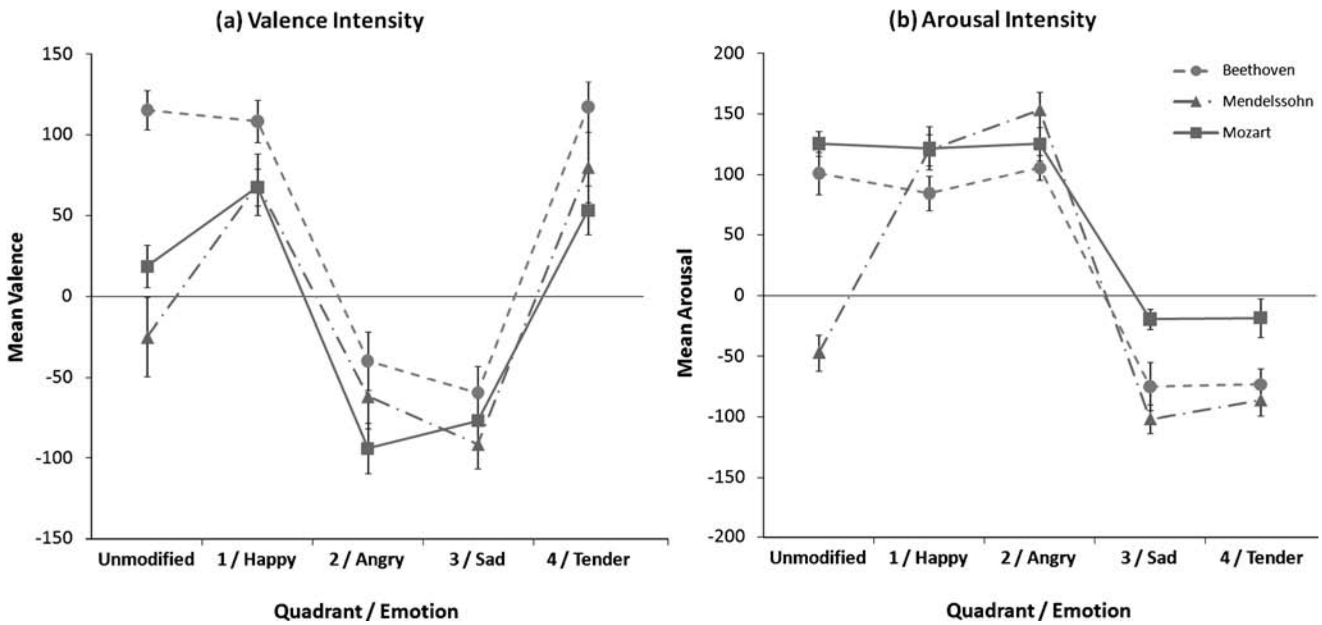
Eighteen university students (9 women, 9 men), 17–30 years old ($\mu = 19.56$; $\sigma = 2.87$) volunteered to participate and were each awarded one-hour of course credit. Nine participants possessed formal music training ($\mu = 2.3$ years; $\sigma = 4.36$ years).

Musical Stimuli

The musical stimuli for CMERS (works 1 and 2) consisted of the same Beethoven and Mendelssohn works used in Experiment 1. For DM, the music (works 3 and 4) consisted of a Mazurka by Cope (1992) in the style of Chopin (unmodified, with a duration of 42 sec); and *Tegnér's Ekorrn*, a Swedish nursery rhyme (unmodified, with a duration of 19 sec). Works 1 and 2 were modified by CMERS only and were taken from Experiment 1. Works 3 and 4 were modified by DM only, and had been used by Bresin and Friberg (2000a) in previous testing. DM-generated files were encoded as 64 kHz Sun audio files and downloaded from www.speech.kth.se/~roberto/emotion.

The unmodified works used by the two systems were matched on emotion. The Beethoven and *Tegnér* works, both composed in the major mode, represent the happy exemplars for the two systems; the Mendelssohn and Cope works, both composed in the minor mode, represent the sad exemplars. All files were generated by the respective authors of CMERS and DM. This choice was done to ensure optimal rule parameters were used for all stimuli. Comparison of the emotionally unmodified works used for the two systems (repeated-measures *t*-tests) revealed no significant difference in participant ratings for either arousal or valence, for either the happy exemplars (Beethoven for CMERS and *Tegnér* for DM) or the sad exemplars (Mendelssohn for CMERS and Cope for DM).

Figure 6. Changes in (a) valence intensity and (b) arousal intensity in Experiment 1.



Testing Apparatus and Participant Handouts

Continuous measurement of user self-reporting was used, with the same tool from Experiment 1. Handouts and the movie tutorial from Experiment 1 were reused.

Procedure

Participants were tested individually with Sennheiser HD 515 headphones, adjusted to a comfortable volume level. Six testing sessions were conducted. Participants heard 20 music samples: 10 from CMERS stimuli works 1 and 2 (four emotionally modified along with the unmodified), and 10 from DM stimuli works 3 and 4 (four emotionally modified along with unmodified). Testing took approximately 60 minutes.

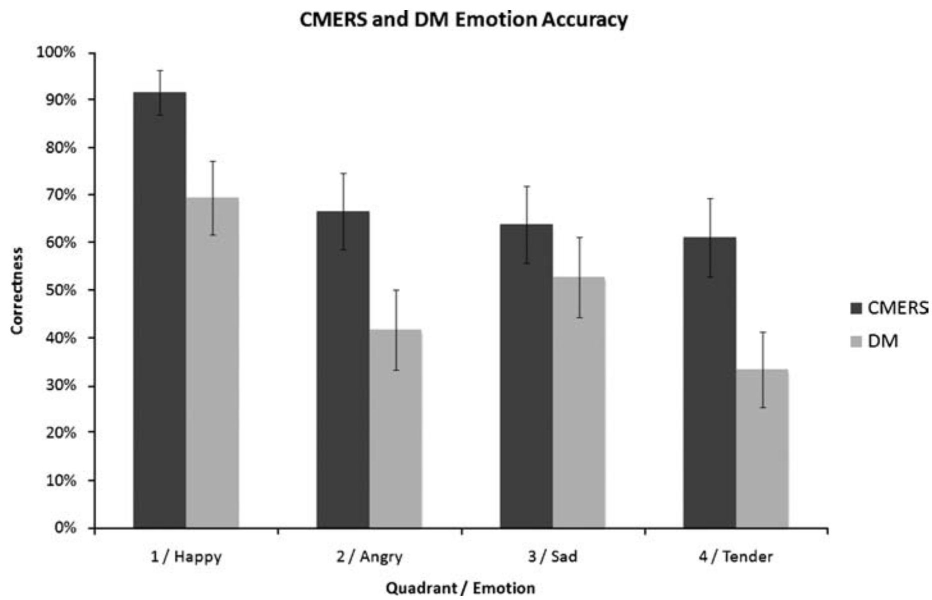
Data Analysis

For hypothesis 1 (that CMERS is more accurate than DM at shifting perceived emotion), coordinate values were reduced to a binary measure of correctness: a rating of 1 if Perceived Emotion matched Intended Emotion, or 0 otherwise. A binary logistic regression was conducted to determine if a significant relationship existed between Correctness (two categories), Quadrant (four categories), System (two categories, i.e., CMERS and DM), and Music Work (two categories, nested under system). Quadrant-by-System interaction was included in the analysis. This interaction was found to be non-significant, and a second analysis was conducted that did not include the interaction.

Table 6. Within-Subject Contrasts Illustrate the Change in Emotion for Quadrants Relative to the Emotionally Unmodified Versions across All Three Music Works in Experiment 1

Dimension	Comparison	F	$p \leq$	η^2
Arousal	Q1 VS Unmodified	16.80	.001	0.47
	Q2 VS Unmodified	45.53	.0005	0.70
	Q3 VS Unmodified	206.76	.0005	0.92
	Q4 VS Unmodified	83.69	.0005	0.82
Valence	Q1 VS Unmodified	9.54	.006	0.33
	Q2 VS Unmodified	72.91	.0005	0.79
	Q3 VS Unmodified	66.77	.0005	0.78
	Q4 VS Unmodified	9.30	.007	0.33

Figure 7. CMERS and DM mean correctness for each emotion across music works in Experiment 2.



For hypotheses 2 and 3 (that CMERS is more effective than DM at shifting valence and arousal), a repeated-measures analysis of variance was conducted with the same independent variables as the binary logistic regression. Correct changes in arousal and valence (as a proportion of total possible change on the 2DES plane) were used as the dependent variables.

A Pearson product-moment correlation coefficient was conducted to analyze the relationship between Play Length (years of playing an instrument), Lesson Years (years of taking music lessons), and CMERS and DM Accuracy.

Results

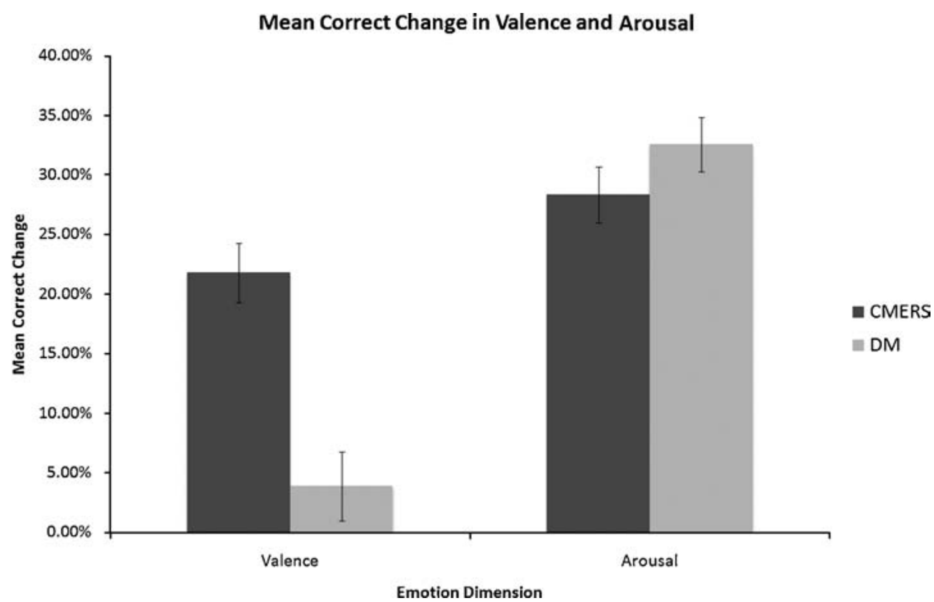
The binary logistic regression analysis revealed an odds ratio of OR = 4.94, $z = 2.80$, $p = .005$. This result suggests that the odds of Correctness with CMERS were approximately five times greater than that of DM. The interaction between Quadrant and System was not significant with $\chi^2(3) = 1.98$, $p = .577$, indicating that CMERS's improved rate of accuracy was not the result of any single quadrant, and that it was consistently more accurate across all quadrants. As this interaction was not significant, it was removed in the second regression analysis.

The second binary logistic regression analysis revealed that the interaction between Music Work and Quadrant on Correctness was significant with $\chi^2(3) = 21.86$, $p = .0001$, indicating that the choice of Quadrant did affect Correctness. Correctness was significantly higher for Quadrant 1, but it was not significant for Quadrants 2, 3, and 4, indicating the interaction effect was owing to Quadrant 1. The effect of Music Work on Correctness was not significant with $\chi^2(2) = 4.08$, $p = .1299$.

The emotion (Quadrant) accuracy means for each system are illustrated in Figure 7. The mean accuracies across the four emotion categories, and across two music works per system, were 71% (CMERS) and 49% (DM).

The repeated-measures analysis of variance for valence found a significant difference between Systems with $F(1, 17) = 45.49$, $p < .0005$. The interaction between System and Quadrant was significant with $F(3, 51) = 4.23$, $p = .01$. These results indicate that CMERS was significantly more effective at correctly influencing valence than DM, with some quadrants more effective than others. For arousal, System was not significant, with $F(1, 17) = 3.54$, $p = .077$. The interaction between System and Quadrant was not significant with $F(3, 51) = 1.62$, $p = .197$. These results indicate that there was no

Figure 8. Mean correct change in valence and arousal for CMERS and DM in Experiment 2.



significant difference between CMERS and DM in arousal modification. The mean correct change in valence and arousal intensities, as a proportion of total possible change on the 2DES plane, for CMERS and DM are illustrated in Figure 8.

The analysis of musical experience found a significant positive correlation between Play Length and CMERS Accuracy ($r = 0.51$, $n = 18$, $p < .05$) but no significant correlation between Play Length and DM Accuracy ($r = 0.134$, $n = 19$, $p = .596$). Lesson Years was not significant for either CMERS or DM Accuracy. No significant correlation was found between CMERS Accuracy and DM Accuracy ($r = 0.095$, $n = 18$, $p = .709$). These results indicate that participants with more instrumental experience were more accurate at detecting the intended emotion of CMERS, but this was not true for DM. Also, participants who were accurate for CMERS were not the same participants who were accurate for DM.

General Discussion

The results of Experiment 1 supported the main hypothesis that CMERS could change perceived judgments of emotion for all selected music works to all four emotion quadrants of the 2DES: happy, angry, sad, and tender. CMERS shifted the perceived emotion with a mean accuracy rate of 78% across musical works and quadrants. This result supported the collective use of CMERS's music-emotion rules in Table 4. System accuracy was not affected by the choice of music work, the intended emotion, or combination of the two, suggesting that CMERS successfully shifted the perceived emotion to all quadrants, regardless of the work's original emotion. This is a significant improvement over an earlier system prototype that achieved a mean accuracy rate of 63% and had a deficit for "anger" (Livingstone and Brown 2005; Livingstone 2008). As the prototype only possessed structural music-emotion rules, the addition of performance music-emotion rules and expressive-feature rules appears to have improved system performance.

The analysis of valence-arousal intensities found that CMERS achieved significant shifts in both dimensions (see Figure 6). This result demonstrated that CMERS could change both the category and intensity of the perceived emotion. This result is further supported by the contrast effects listed in Table 6, with CMERS shifting the intensity of valence and arousal for all works, regardless of their original emotion. Table 6 also indicated that the degree of emotional shift was constrained by the original emotion of the unmodified work. That is, for the two works classified as "happy" (high valence and arousal, Quadrant 1), CMERS could only slightly increase their valence and arousal (i.e., make them "happier"). Conversely, the most significant shift occurred for Quadrant 3, the polar opposite of happy.

Results of Experiment 2 supported the main hypothesis that CMERS was significantly more accurate at influencing judgments of perceived musical emotion than was DM. CMERS achieved a mean accuracy of 71% and DM of 49% across the four emotions. These results support the claim that a system operating on both score and performance data is significantly more successful in changing perceived emotion than a system focusing on performance data only. Of particular interest were the systems' respective capabilities in influencing valence and arousal (see Figure 8). Although both systems performed comparably on arousal, CMERS was the only system capable of creating a significant change in valence. As previously discussed, only two music-emotion rule types can potentially shift the valence of a work: Mode and Harmonic complexity (see Figure 2). As CMERS modified the mode, and DM did not modify either feature, it is thought that this difference enabled CMERS to achieve significant shifts in valence. Future testing of CMERS with isolated feature modification will examine this.

As CMERS and DM were tested using separate stimuli, differences in system performances could be attributed to this methodology. However, as previously discussed, a statistical comparison of the emotionally unmodified works used for the two systems revealed no significant differences in either arousal or valence. Therefore, any differences in the music samples cannot explain differential effects of the two systems in changing emotional quality in both directions along the two dimensions. This is supported by Figure 8, which illustrates that CMERS and DM both made significant changes to arousal. If differences were the result of different stimuli, this effect would not have been observed. Rather, both valence and arousal would have been adversely affected. As discussed, we attribute this difference in Figure 8 to a fundamental difference in the rule sets used by the two systems.

Prior testing of DM by Bresin and Friberg (2000a), which used the same music files as Experiment 2 in this article, reported that listeners correctly identified the intended emotion 63% of the time across seven emotional categories. However, Experiment 2 reported DM emotion identification accuracy at 49% across four emotional categories. This difference may be the result of differing response methodologies. Prior testing of DM used a single forced-choice rank-and-match measure (Schubert 1999a). This methodology had the user match seven music samples to seven emotion categories. With this methodology, users may have been able to differentiate the attempted modifications; however, the intensity of the emotion was not evaluated. Thus, it was not gauged whether the music samples achieved the desired emotion—only that the modifications were reliably differentiable. Continuous response with a dimensional representation may therefore offer a more accurate approach in music-emotion research.

Finally, Experiment 2 found that participants were more accurate at identifying CMERS's intended emotion when they possessed more years of instrumental experience. This result was not found for DM. Interestingly, participants who were accurate at CMERS were not the same participants who were accurate at DM. Together, these results suggest that CMERS manipulates particular features of the music that can only be accurately decoded after years of musical instrument experience. These features are learned, the presence of which benefits trained individuals. This result adds to a body of literature on the relationship between training and music cognition (Bigand and Poulin-Charronnat 2006). Future testing with trained and untrained participants will help identify how the perception of features varies with musical experience, and which features are important for communicating emotion for each group.

Two philosophical questions are often raised in the context of CMERS. The first is whether changes to a work violate the composer's original intentions, decreasing its "musicality." The goal of CMERS is to provide a tool to investigate the precise effects of musical features on emotion. Musicality is a separate line of research that requires a different response methodology (e.g., listener ratings of musical quality or preference). This would locate feature modification limits, beyond which musicality is adversely affected. Such testing would be required before CMERS and similar systems could be deployed in a real-world context (e.g., computer games). The second question involves the degree to which a musical work can be modified and still be considered the same work. This is a difficult question to answer, considering the extent of musical variations sometimes created by composers and improvisers. For example, can Jimi Hendrix's iconoclastic performance of *The Star-Spangled Banner* at Woodstock in 1969 still be considered the same work as the original anthem? Again, this is a separate line of research and could involve perceptual evaluations using a same/different task with a set of musical variations.

Conclusions and Outlook

Although CMERS achieved its experimental objectives, there are avenues for further research. Foremost, the rule set could be expanded to include the complete set of Primary Music-Emotion Rules. The set of expressive performance features could also be expanded, with a refinement of implementation details and parameter values. The prioritization of rules in CMERS was based on empirical citation counts, not perceptual effect sizes. The goal of this research project is now to investigate the perceptual effects of these features, and determine their relative influence. Improvements to data-analysis techniques should also be incorporated. Although continuous feedback was captured in both rounds of testing, this data was averaged to a single arousal-valence coordinate value per music sample. Future testing will use Functional Data Analysis to examine the temporal nature of participant responses and how responses correlate to modified features and the work's phrase structure (Livingstone et al. 2009). This technique will also enable an examination of CMERS's real-time modification capability, which was not examined in this article.

The current system is specific to Western classical music, and its effectiveness on other genres is unclear. The CMERS rule set was drawn from empirical studies that focused on Western classical music. However, earlier testing of a system prototype achieved comparable results with a non-Western musical work used in a computer game (Livingstone and Brown 2005; Livingstone 2008). Considering recent cross-cultural findings (Fritz et al. 2009; see also Thompson and Balkwill 2010), this suggests that CMERS may be effective with other musical genres given a modified rule set. CMERS's effects on other aspects of emotion should also be investigated. This study examined perceived emotion, not felt emotion (Gabrielsson 2002). However, listeners report they often feel emotion similar to that perceived (Juslin and Laukka 2004; Evans and Schubert 2006). Musical expectancy and tension is an important aspect of emotion that was not examined in the present article (Meyer 1956; Juslin and Västfjäll 2008). The effect of CMERS on musical expectancy would be an interesting line of future research. An open-source implementation of CMERS in Matlab/Octave is under development and will be made freely available in the future online at www.itee.uq.edu.au/~srl/CMERS.

Computational modeling and systematic manipulation provides researchers with a powerful methodology for exploring the effects of individual music features on emotion. CMERS's capacity to isolate and control individual features without the bias of a performer provides researchers with a tool to further develop theories of music and emotion.

Acknowledgments

The authors wish to thank Andrew Sorensen, Anders Friberg, Alan Taylor, and Pascale Lidji for their assistance, contributions, and comments.

References

Arcos, J. L., et al. 1998. "SaxEx: A Case-Based Reasoning System for Generating Expressive Musical Performance." *Journal of New Music Research* 27(3):194–210.
Bengtsson, I., and A. Gabrielsson. 1980. "Methods for Analyzing Performances of Musical Rhythm." *Scandinavian Journal of Psychology* 21(1):257–268.
Bigand, E., and B. Poulin-Charronnat. 2006. "Are We 'Experienced Listeners?' A Review of the Musical Capacities that do not Depend on Formal Musical Training." *Cognition* 100(1):100–130.

Bresin, R. 1993. "MELODIA: A Program for Performance Rules Testing, Teaching, and Piano Score Performance." *Proceedings of the X Colloquio di Informatica Musicale*. Milano: University of Milano, pp. 325–327.

Bresin, R., and A. Friberg. 2000a. "Emotional Coloring of Computer-Controlled Music Performances." *Computer Music Journal* 24(4):44–63.

Bresin, R., and A. Friberg. 2000b. "Rule-Based Emotional Colouring of Music Performance." *Proceedings of the 2000 International Computer Music Conference*. San Francisco, California: International Computer Music Association, pp. 364–367.

Brunswik, E. 1952. "The Conceptual Framework of Psychology." *International Encyclopedia of Unified Science*. Chicago, Illinois: University of Chicago Press.

Brunswik, E. 1956. *Perception and the Representative Design of Psychological Experiments*, 2nd ed. Berkeley: University of California Press.

Canazza, S., et al. 2004. "Modeling and Control of Expressiveness in Music Performance." *Proceedings of the IEEE* 92(4):686–701.

Chew, G. 2007. "Slur." In *Grove Music Online*. Available online at www.oxfordmusiconline.com/subscriber/article/grove/music/25977.

Clynes, M. 1998. *SuperConductor: The Global Music Interpretation and Performance Program*. Available online at www.microsoundmusic.com/clynes/superc.htm.

Cope, D. 1992. "Computer Modeling of Musical Intelligence in Experiments in Musical Intelligence." *Computer Music Journal* 16(2):69–83.

Ekman, P. 1973. *Darwin and Facial Expression: A Century of Research in Review*. New York: Academic Press.

Evans, P., and E. Schubert. 2006. "Quantification of Gabrielsson's Relationships Between Felt and Expressed Emotions in Music." *Proceedings of the 9th International Conference on Music Perception and Cognition*. Bologna: ESCOM, pp. 446–454.

Friberg, A. 1991. "Generative Rules for Music Performance: A Formal Description of a Rule System." *Computer Music Journal* 15(2):56–71.

Friberg, A. 1995. "Matching the Rule Parameters of Phrase Arch to Performances of Träumerei: A Preliminary Study." In A. Friberg and J. Sundberg, eds. *Proceedings of the KTH Symposium on Grammars for Music Performance*. Stockholm: Royal Institute of Technology, pp. 37–44.

Friberg, A. 2006. "pDM: An Expressive Sequencer with Real-Time Control of the KTH Music-Performance Rules." *Computer Music Journal* 30(1):37–48.

Friberg, A., R. Bresin, and J. Sundberg. 2006. "Overview of the KTH Rule System for Musical Performance." *Advances in Cognitive Psychology* 2(2–3):145–161.

Friberg, A., and J. Sundberg. 1999. "Does Music Performance Allude to Locomotion? A Model of Final Ritardandi Derived from Measurements of Stopping Runners." *Journal of the Acoustical Society of America* 105(3):1469–1484.

Fritz, T., et al. 2009. "Universal Recognition of Three Basic Emotions in Music." *Current Biology* 19(7):573–576.

Gabrielsson, A. 1985. "Interplay Between Analysis and Synthesis in Studies of Music Performance and Music Experience." *Music Perception* 3(1):59–86.

Gabrielsson, A. 1988. "Timing in Music Performance and its Relations to Music Expression." In J. A. Sloboda, ed. *Generative Processes in Music*. Oxford: Oxford University Press, pp. 27–51.

Gabrielsson, A. 1994. "Intention and Emotional Expression in Music Performance." In A. Friberg et al., eds. *Proceedings of the 1993 Stockholm Music Acoustics Conference*. Stockholm: Royal Swedish Academy of Music, pp. 108–111.

Gabrielsson, A. 1995. "Expressive Intention and Performance." In R. Steinberg, ed. *Music and the Mind Machine: The Psychophysiology and the Psychopathology of the Sense of Music*. Berlin: Springer Verlag, pp. 35–47.

Gabrielsson, A. 1999. "The Performance of Music." In D. Deutsch, ed. *The Psychology of Music*, 2nd ed. San Diego, California: Academic Press, pp. 501–602.

Gabrielsson, A. 2002. "Perceived Emotion and Felt Emotion: Same or Different?" *Musicae Scientiae* Special Issue 2001–2002:123–148.

Gabrielsson, A. 2003. "Music Performance Research at the Millennium." *The Psychology of Music* 31(3):221–272.

Gabrielsson, A., and E. Lindstrom. 2001. "The Influence of Musical Structure on Emotional Expression." In P. N. Juslin and J. A. Sloboda, eds. *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, pp. 223–248.

Goebl, W. 2001. "Melody Lead in Piano Performance: Expressive Device or Artifact?" *Journal of the Acoustical Society of America* 110(1):563–572.

Hevner, K. 1935. "The Affective Character of the Major and Minor Modes in Music." *American Journal of Psychology* 47(1):103–118.

Honing, H. 1990. "Poco: An Environment for Analysing, Modifying, and Generating Expression in Music." *Proceedings of the 1990 International Computer Music Conference*. San Francisco, California: International Computer Music Association, pp. 364–368.

Ilie, G., and W. F. Thompson. 2006. "A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect." *Music Perception* 23(4):319–330.

Juslin, P. N. 1997a. "Can Results from Studies of Perceived Expression in Musical Performances be Generalized Across Response Formats?" *Psychomusicology* 16(1–2):77–101.

Juslin, P. N. 1997b. "Emotional Communication in Music Performance: A Functionalist Perspective and Some Data." *Music Perception* 14(4):383–418.

Juslin, P. N. 1997c. "Perceived Emotional Expression in Synthesized Performances of a Short Melody: Capturing the Listener's Judgment Policy." *Musicae Scientiae* 1:225–256.

Juslin, P. N. 2001. "Communicating Emotion in Music Performance: A Review and a Theoretical Framework." In P. N. Juslin and J. A. Sloboda, eds. *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, pp. 309–340.

Juslin, P. N., and P. Laukka. 2003. "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?" *Psychological Bulletin* 129(5):770–814.

Juslin, P. N., and P. Laukka. 2004. "Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening." *Journal of New Music Research* 33(3):217–238.

Juslin, P. N., and D. Västfjäll. 2008. "Emotional Responses to Music: The Need to Consider Underlying Mechanisms." *Behavioral and Brain Sciences* 31(5):559–575.

Kastner, M. P., and R. G. Crowder. 1990. "Perception of the Major/Minor Distinction: IV. Emotional Connotations in Young Children." *Music Perception* 8(2):189–202.

Kendall, R. A., and E. C. Carterette. 1990. "The Communication of Musical Expression." *Music Perception* 8(2):129–164.

Kronman, U., and J. Sundberg. 1987. "Is the Musical Retard an Allusion to Physical Motion?" In A. Gabrielsson, ed. *Action and Perception in Rhythm and Music*. Stockholm: Royal Swedish Academy of Music, Publication No. 55, pp. 57–68.

Krumhansl, C. L., and E. J. Kessler. 1982. "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys." *Psychological Review* 89(4):334–368.

Lerdahl, F., and R. Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, Massachusetts: MIT Press.

Lindström, E. 2004. "A Dynamic View of Melodic Organization and Performance: Perception of Structure and Emotional Expression in Music." In *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences* 138. Uppsala: Uppsala University Library. Available online at http://uu.diva-portal.org/smash/record.jsf?pid=diva2:164510.

Livingstone, S. R. 2008. "Changing Musical Emotion through Score and Performance with a Computational Rule System." PhD thesis, School of Information Technology and Electrical Engineering, University of Queensland.

Livingstone, S. R., and A. R. Brown. 2005. "Dynamic Response: Real-Time Adaptation for Music Emotion." *Proceedings of the Second Australasian Conference on Interactive Entertainment*. Sydney: Creativity and Cognition Studios Press, pp. 105–111.

Livingstone, S. R., A. R. Brown, and R. Muhlberger. 2005. "Influencing the Perceived Emotions of Music with Intent." In T. Innocent, ed. *Third Iteration: Proceedings of the Third International Conference on Generative Systems in the Electronic Arts*. Victoria: Centre for Electronic Media Art, pp. 161–170.

Livingstone, S. R., and W. F. Thompson. 2006. "Multimodal Affective Interaction: A Comment on Musical Origins." *Music Perception* 24(1):89–94.

Livingstone, S. R., and W. F. Thompson. 2009. "The Emergence of Music from the Theory of Mind." *Musicae Scientiae* Special Issue 2009/10:83–115.

Livingstone, S. R., W. F. Thompson, and F. A. Russo. 2009. "Facial Expressions and Emotional Singing: A Study of Perception and Production with Motion Capture and Electromyography." *Music Perception* 26(5):475–488.

Livingstone, S. R., et al. 2007. "Controlling Musical Emotionality: An Affective Computational Architecture for Influencing Musical Emotions." *Digital Creativity* 18(1):43–53.

Livingstone, S. R., et al. 2009. "Emotional Arousal and the Automatic Detection of Musical Phrase Boundaries." In A. Williamon, S. Pretty, and R. Buck, eds. *Proceedings of the International Symposium on Performance Science 2009*. Utrecht: European Association of Conservatoires (AEC), pp. 445–450.

Mathews, M. V., and F. R. Moore. 1970. "GROOVE—A Program to Compose, Store, and Edit Functions of Time." *Communications of the ACM* 13(12):715–721.

Mazzola, G., and O. Zahorka. 1994. "The RUBATO Performance Workstation on NEXTSTEP." *Proceedings of the 1994 International Computer Music Conference*. San Francisco, California: International Computer Music Association, pp. 102–108.

Meyer, Leonard B. 1956. *Emotion and Meaning in Music*. Chicago, Illinois: University of Chicago Press.

Mion, L., et al. 2010. "Modeling Expression with Perceptual Audio Features to Enhance User Interaction." *Computer Music Journal* 34(1):65–79.

Oliveira, A. P., and A. Cardoso. 2009. "Automatic Manipulation of Music to Express Desired Emotions." In F. Gouyon, A. Barbosa, and X. Serra, eds. *Proceedings of the 6th Sound and Music Computing Conference*. Porto: University of Porto, pp. 265–270.

Palmer, C. 1989. "Mapping Musical Thought to Musical Performance." *Journal of Experimental Psychology: Human Perception and Performance* 15(2):331–346.

Palmer, C. 1992. "The Role of Interpretive Preferences in Music Performance." In M. R. Jones and S. Holleran, eds. *Cognitive Bases of Musical Communication*. Washington, DC: American Psychological Association, pp. 249–262.

Palmer, C. 1997. "Music Performance." *Annual Review of Psychology* 48(1):115–138.

Peretz, I., L. Gagnon, and B. Bouchard. 1998. "Music and Emotion: Perceptual Determinants, Immediacy, and Isolation after Brain Damage." *Cognition* 68(2):111–141.

Ramirez, R., and A. Hazan. 2005. "Modeling Expressive Performance in Jazz." *Proceedings of 18th International Florida Artificial Intelligence Research Society Conference.* Menlo Park, California: Association for the Advancement of Artificial Intelligence Press, pp. 86–91.

Repp, B. H. 1992. "Diversity and Commonality in Music Performance: An Analysis of Timing Microstructure in Schumann's 'Träumerei'." *Journal of the Acoustical Society of America* 92(5):2546–2568.

Repp, B. H. 1999. "A Microcosm of Musical Expression: II. Quantitative Analysis of Pianists' Dynamics in the Initial Measures of Chopin's Etude in E major." *Journal of the Acoustical Society of America* 105(3):1972–1988.

Russell, J. A. 1980. "A Circumplex Model of Affect." *Journal of Social Psychology* 39(6):1161–1178.

Scherer, K. R. 1986. "Vocal Affect Expression: A Review and a Model for Future Research." *Psychological Bulletin* 99(2):143–165.

Schubert, E. 1999a. "Measurement and Time Series Analysis of Emotion in Music." PhD thesis, University of New South Wales.

Schubert, E. 1999b. "Measuring Emotion Continuously: Validity and Reliability of the Two-Dimensional Emotion-Space." *Australian Journal of Psychology* 51(3):154–165.

Schubert, E. 2001. "Continuous Measurement of Self-Report Emotional Response to Music." In P. N. Juslin and J. A. Sloboda, eds. *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, pp. 393–414.

Schubert, E. 2004. "Modeling Perceived Emotion with Continuous Musical Features." *Music Perception* 21(4):561–585.

Seashore, C. E. 1923. "Measurement of the Expression of Emotion in Music." *Proceedings of the National Academy of Sciences* 9(9):323–325.

Sorensen, A. C., and A. R. Brown. 2007. "aa-cell in Practice: An Approach to Musical Live Coding." *Proceedings of the 2007 International Computer Music Conference*. San Francisco, California: International Computer Music Association, pp. 292–299.

Sundberg, J., A. Askenfelt, and L. Frydén. 1983. "Musical Performance: A Synthesis-by-Rule Approach." *Computer Music Journal* 7(1):37–43.

Sundberg, J., A. Friberg, and R. Bresin. 2003. "Musician's and Computer's Tone Inter-Onset-Interval in Mozart's Piano Sonata K 332, 2nd mvt, bar 1-20." *TMH-QPSR* 45(1):47–59.

Temperley, D. 2004. "An Evaluation System for Metrical Models." *Computer Music Journal* 28(3):28–44.

Thompson, W. F., and L.-L. Balkwill. 2010. "Cross-Cultural Similarities and Differences." In P. N. Juslin and J. A. Sloboda, eds. *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford: Oxford University Press, pp. 755–788.

Thompson, W. F., and B. Robitaille. 1992. "Can Composers Express Emotions Through Music?" *Empirical Studies of the Arts* 10(1):79–89.

Todd, N. P. M. 1985. "A Model of Expressive Timing in Tonal Music." *Music Perception* 3(1):33–58.

Todd, N. P. M. 1989a. "A Computational Model of Rubato." *Contemporary Music Review* 3(1):69–88.

Todd, N. P. M. 1989b. "Towards a Cognitive Theory of Expression: The Performance and Perception of Rubato." *Contemporary Music Review* 4(1):405–416.

Todd, N. P. M. 1992. "The Dynamics of Dynamics: A Model of Musical Expression." *Journal of the Acoustical Society of America* 91(6):3540–3550.

Widmer, G., and A. Tobudic. 2003. "Playing Mozart by Analogy: Learning Multi-Level Timing and Dynamics Strategies." *Journal of New Music Research* 32(3):259–268.

Winter, R. 2006. "Interactive Music: Compositional Techniques for Communicating Different Emotional Qualities." Master's thesis, University of York.