Play It Again With Feeling: Computer Feedback in Musical Communication of Emotions

Patrik N. Juslin, Jessika Karlsson, Erik Lindström, Anders Friberg, and Erwin Schoonderwaldt
Uppsala University

Communication of emotions is of crucial importance in music performance. Yet research has suggested that this skill is neglected in music education. This article presents and evaluates a computer program that automatically analyzes music performances and provides feedback to musicians in order to enhance their communication of emotions. Thirty-six semiprofessional jazz/rock guitar players were randomly assigned to one of 3 conditions: (1) feedback from the computer program, (2) feedback from music teachers, and (3) repetition without feedback. Performance measures revealed the greatest improvement in communication accuracy for the computer program, but usability measures indicated that certain aspects of the program could be improved. Implications for music education are discussed.

Keywords: music performance, communication, emotion, computer-based teaching, feedback

The most profound moments of musical experience often derive from a performer's ability to communicate heartfelt emotions to the listener. Yet, emotional aspects are often neglected in music education, perhaps because communication of emotions involves tacit knowledge that is difficult to convey from teacher to student. This article presents a new, empirically based approach to learning communication of emotion that involves feedback from a computer program. First, we briefly summarize previous research and outline the program. Then, we report three experiments that explored the efficacy and usability of the program. Finally, we discuss implications of the new approach for music education.

Previous Research

Musical Expressivity

One of the primary themes in the study of music and its performance is that music is heard as expressive by listeners (Budd, 1985; Davies, 1994; Ratner, 1980). People become moved by particularly expressive performances, which for many listeners is the essence of music. Moreover, questionnaire research suggests that performers and music teachers view expression as the most crucial aspect of a performer's skills (e.g., Laukka, 2004; Lind-

ström, Juslin, Bresin, & Williamon, 2003). Clearly, good technique is required to master a musical instrument, but expression is what really sets performers apart (see Boyd & George-Warren, 1992, pp. 103–108).

Yet, the nature of expressivity itself has largely been shrouded in mystery. Only in the last decade has empirical research yielded a better understanding of the nature of expressive performance. Following the lead of Seashore's (1938) seminal work, we will use *expression* to refer to the psychophysical relationships among objective characteristics of the music and subjective impressions of the listener. More recent research has indicated that expression is a multidimensional phenomenon (Juslin, 2003; Juslin, Friberg, & Bresin, 2002) consisting of distinct components of information that involve marking of musical structure (Clarke, 1988), expression of specific emotions (Juslin, 1997a), and giving the music an appropriate motion character (Shove & Repp, 1995). In this article, we will focus on the emotion component of expressivity, while acknowledging that this is not the only important aspect, because clearly it is the emotion component that is most strongly associated with the notion of expression in music (Budd, 1985; Gabrielsson & Juslin, 2003; Juslin & Laukka, 2004; Matthay, 1913).

Music as Communication of Emotions

Emotional expression in music performance is commonly conceptualized in terms of a communication process, in which musicians *encode* (or express) particular emotions that are *decoded* (or recognized) by listeners (Juslin, 2005; Thompson & Robitaille, 1992). Although some authors have objected to this notion (Budd, 1989; Serafine, 1980), evidence supporting the notion comes from two kinds of sources.

First, 45 studies have provided compelling evidence that professional performers are able to communicate discrete emotions to listeners by using acoustic features, such as tempo, sound level, articulation, and timbre (for a review, see Juslin & Laukka, 2003). The accuracy with which the emotions are communicated approaches that of facial and vocal expression of emotions. Most of these studies have used a procedure in which musicians were asked

to play short pieces of music in order to express different emotions (e.g., sadness). The performances were recorded and used in listening tests to see whether listeners could accurately decode the intended expression. Many studies also analyzed the acoustic features of the performances to explore how each emotion was expressed. Such analyses have produced detailed descriptions of the acoustic features used to express various emotions (Juslin, 2001, Figure 14.2).

Second, further evidence supporting the notion of music-as-communication comes from questionnaire studies and interviews with musicians and listeners. In a study featuring 145 listeners (aged 17–74), the majority of the participants reported experiencing that music communicates emotions, as revealed by their own free responses to an open-ended question, and 76% of them responded that music expresses emotions "often" (Juslin & Laukka, 2004). Similarly, a questionnaire study featuring 135 expert musicians revealed that the majority of the musicians defined expression mainly in terms of "communicating emotions" and "playing with feeling," as indicated by their own free responses (Lindström et al., 2003). Furthermore, 83% of the musicians claimed that they try to express specific emotions in their performance "always" or "often." Minassian, Gayford, and Sloboda (2003) conducted a questionnaire study featuring 53 high-level classical performers, and investigated which factors were statistically associated with an "optimal" performance. Performances judged as optimal tended to be those where the performer (a) had a clear intention to communicate (usually an emotional message), (b) was emotionally engaged with the music, and (c) believed the message had been received by the audience. Hence, it seems safe to conclude that communication of emotion is a crucial aspect of music performance that a musician needs to address in order to be successful.

Emotion in Music Education

In view of these findings, one would expect expressive skills to be given high priority by music teachers. Although this indeed seems to be the case (Laukka, 2004), many studies have suggested that music teaching focuses mainly on technique rather than on expressivity (Hepler, 1986; Persson, 1993; Rostvall & West, 2001; Tait, 1992), and many method books for music instrument teaching do not cover expressive aspects at all (e.g., Rostwall & West, 2001). This neglect of expressivity may result in students developing expressive skills rather late in their artistic development. Thus, for example, 48% of the music students in Woody's (2000) questionnaire study did not become "seriously concerned with expressivity" until they were well into high school, or even in their first year of college.

Closer examination of the literature on music education reveals that this concern is not exactly new: More than 40 years ago, Hoffren (1964) observed that expression was a "neglected area" reflecting "the present American obsession with technique" (p. 32); Marchand (1975) voiced "a suspicion that performance teaching/learning is too technique-oriented" and that programs solely devoted to technical skills may yield "performers who lack expression in their playing" (p. 14); Reimer (2003) encouraged music educators to devote more attention to emotion and expression in music, arguing that the emotional dimension of music "is probably its most important defining characteristic" (p. 72). Still, little had apparently changed when Juslin and Persson (2002) reviewed the

topic nearly 40 years after the first critical remarks. Why has expression continued to be neglected in music education?

First, the nature of expression does not lend itself easily to formalized description; for instance, much knowledge about expression is tacit and therefore difficult to express in words (e.g., Hoffren, 1964). This is problematic because teaching is apparently dominated by verbal instruction (Karlsson & Juslin, 2005). Second, studies of how performers express emotions in music performance only matured in the last decade (Juslin & Laukka, 2003, Figure 1). Hence, researchers have not been able to provide teachers with theories or findings that could guide teaching. Instead of providing explicit instruction with respect to emotional expression, teachers have mostly used strategies that address expression only indirectly.

Traditional Teaching Strategies

One of the traditional strategies used to teach a student how a piece of music should be performed is *musical modeling* (Dickey, 1992). The teacher's performance provides a model of what is desired from the student and the student is required to learn by imitating this model. Although modeling is useful (e.g., Ebie, 2004), it has certain limitations. One limitation is that the student is required to pick up the relevant aspects of the model and that it can be hard for a student to know what to listen for and how to represent it in terms of specific skills (Lehmann, 1997). Furthermore, some authors worry that imitation might lead to superficial skills that are difficult to apply to new situations (Tait, 1992).

However, there are a number of experiential teaching strategies also, which instead aim at conveying the subjective aspects of performing to a student. One such strategy is the use of *metaphors*. Metaphors are used to focus the emotional qualities of the performance by serving as a reference or evoking a mood within the performer (Barten, 1998; Rosenberg & Trusheim, 1989). For example, a teacher may say: "Close your eyes and think about how you would feel if you received a phone call later that day saying a close friend or relative was just killed in an accident" (Bruser, 1997, p. 57). Although metaphors can be effective, there are problems with them. For instance, metaphors depend on the performer's personal experience with words and images, and because different performers have different experiences, metaphors are frequently ambiguous (e.g., Persson, 1996, pp. 310–311).

Another teaching strategy endorsed by some teachers is to focus on the performer's *felt emotions* (Woody, 2000), trusting that these emotions will naturally translate into appropriate sound properties. Many music students and teachers believe that the emotions must be felt by the performer in order to be communicated well (e.g., Laukka, 2004; Lindström et al., 2003). However, felt emotion provides no guarantee that the emotion will be successfully conveyed to listeners, neither is it necessary to feel the emotion in order to communicate it successfully. On the contrary, strong emotional involvement could lead to muscle tension, with detrimental effects on the performance (Gellrich, 1991).

Finally, teachers may use *musical directions*; that is, comments that directly address the relevant acoustic parameters. Woody (1999), for example, has argued that "the most effective approach for expressive performance involves conscious identification and implementation of specific expressive features" (p. 339). However,

to be successful this strategy requires that the teacher has explicit knowledge about expression, which may not always be the case.

The Importance of Feedback

What is required for effective learning to occur? Based on their extensive overview of a century of research on skill acquisition, Ericsson, Krampe, and Tesch-Römer (1993) proposed three elements that are required in a learning task for it to qualify as *deliberate practice*: (a) a well-defined task, (b) informative feedback, and (c) opportunities for repetition and correction of errors. Feedback is defined as a "process by which an environment returns to individuals a portion of the information in their response output necessary to compare their present strategy with a representation of an ideal strategy" (Balzer, Doherty, & O'Connor, 1989, p. 412). This definition suggests that many traditional teaching strategies (e.g., metaphors and felt emotion) do not provide informative feedback, because they do not provide the performer with a direct comparison of his or her current performance strategy with an optimal strategy. In his review, Tait (1992, p. 532) concluded that "teaching strategies need to become more specific in terms of tasks and feedback." Can empirical research help to solve this problem?

In fact, a number of recent projects have aimed to explore novel approaches to teaching expressivity (Dalgarno, 1997; Johnson, 1998; Sloboda, Minassian, & Gayford, 2003; Woody, 1999), but none of these have focused specifically on communication of emotions. Therefore, in a project called *Feedback Learning of Musical Expressivity* (*Feel-ME*), the present authors have developed a novel computer program that aims to enhance a performer's communication of emotions by providing feedback according to the above definition (Juslin, Laukka, Friberg, Bresin, & Lindström, 2001). The program is intended as a complement to traditional teaching strategies aimed at enhancing expressivity.

A New Approach

Brunswik's Lens Model

Communication of emotion requires that there is both a performer's intention to express an emotion and recognition of this same emotion by a listener. Such communication involves the ability to vary in appropriate ways a number of different musical features: fast-slow, loud-soft, staccato-legato, bright-dark, and so forth. These features have certain characteristics that are crucial to understand in order to devise an efficient teaching strategy. Juslin (1995, 2000) suggested that we should use a variant of Brunswik's (1956) *lens model* to capture the special characteristics of the communicative process, and this model forms the basis of the computer program that we have developed.

The Brunswikian lens model (see Figure 1) can be used to illustrate how musicians are able to communicate emotions by employing a set of acoustic cues (bits of information) such as
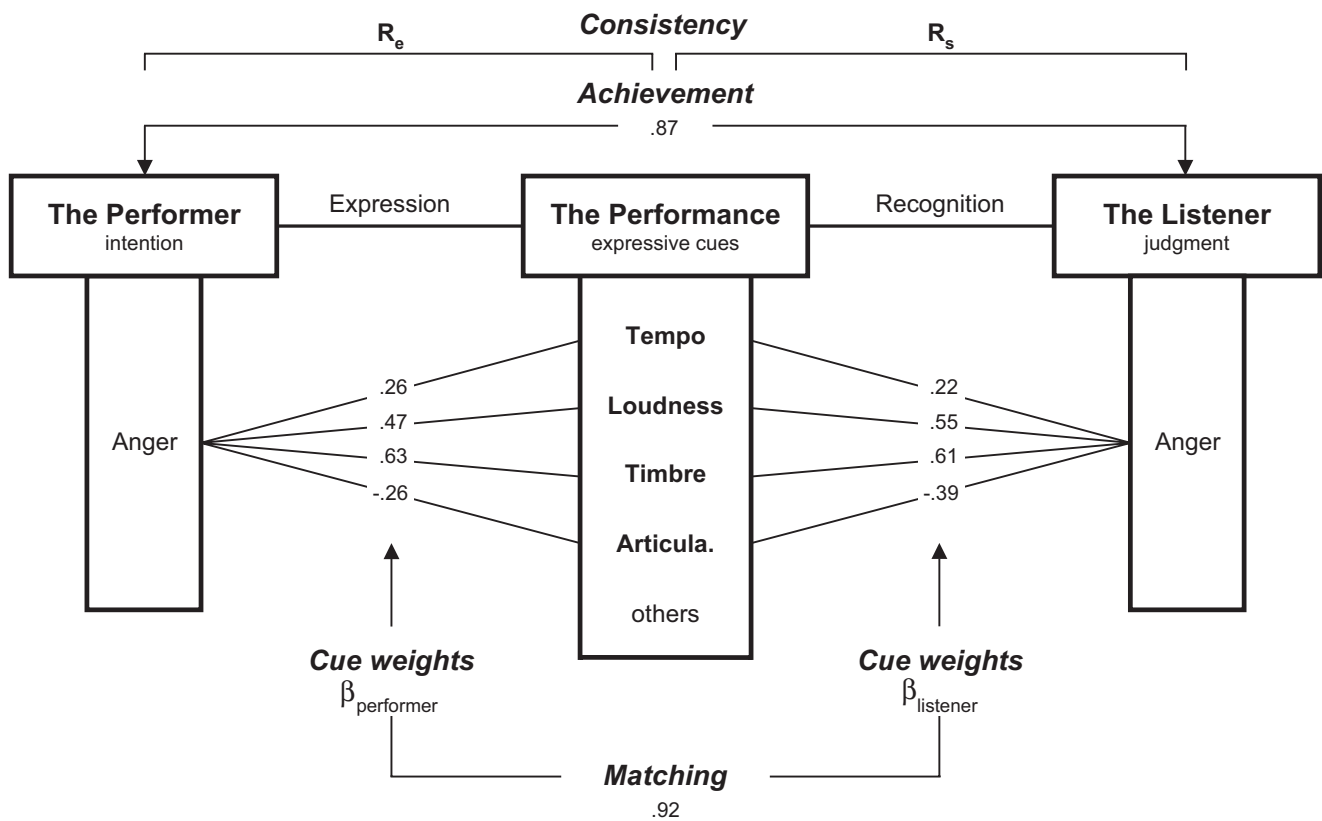

Figure 1. A modified version of Brunswik's lens (1956) model for communication of emotions in music performance (adapted from Juslin, 2000).

tempo, sound level, and timbre that are uncertain, but partly redundant (the cues covary to some extent). The expressed emotions are recognized by listeners, who use these same cues to recognize the intended expression. The cues are uncertain since they are not perfectly reliable indicators of the intended expression. Thus, for instance, fast tempo is not perfectly correlated with expression of happiness, because fast tempo is also used in expression of anger. None of the cues, therefore, is completely reliable when used in isolation, but by combining the values of a number of cues, performers and listeners can achieve reliable expression and recognition, respectively. Listeners integrate the various cues in a mainly additive fashion in their emotion judgments, which can explain how the communication can be successful on different musical instruments that offer different cues. Brunswik's notion of *vicarious functioning* may be used to describe how listeners use partly interchangeable cues in flexible ways, sometimes shifting from a cue that is unavailable to one that is available. Cues are interchangeable, because they are partly redundant. The redundancy among cues partly reflects how sounds are produced on musical instruments (e.g., a harder string attack produces a tone, both louder and sharper in timbre). (For further evidence and discussion of the lens model, see Juslin, 2000, and Juslin & Laukka, 2003.) The relationships between the performer's intention, the acoustic cues, and the listener's judgment can be modeled using correlational statistics. There are several indices in the lens model that are key in understanding the communicative process. (For a description of how each index is measured, see Method section.)

> *Achievement*(r_a) refers to the relationship between the performer's expressive intention (e.g., intending to express sadness) and the listener's judgment (e.g., perceiving sadness). It is a measure of how well the performer succeeds in communicating a given emotion to listeners.

> *Cue weight* (β_1, β_2, β_3 . . .) refers to the strength of the relationship between an individual cue (e.g., tempo), on the one hand, and a performer's intentions or listeners' judgments, on the other. Cue weights indicate how individual cues are actually used by performers and listeners, respectively (e.g., that the performer uses fast tempo to express anger or that listeners use fast tempo to recognize anger).

> *Matching* (G) refers to the degree of similarity between the performer's and the listeners' use of acoustic cues, respectively. For effective communication to occur, the performer's use of cues (i.e., his or her cue weights) must be reasonably matched to the listeners' use of cues.

> *Consistency* (R_e and R_s) refers to the degree of consistency with which the performer and listeners, respectively, are able to use the cues. Other things equal, the communication will be more effective if the cues are used consistently.

The relations among the different indices of the lens model have been mathematically formulated in terms of the *lens model equation* (Hursch, Hammond, & Hursch, 1964), which allows one to explain achievement in terms of matching and consis-

tency. The essential point in the present context is that the upper limit of achievement is set by the matching, performer consistency, and listener consistency. Therefore, if the musical communication of an emotion is unsuccessful, this could be because (1) performer and listeners use the cues differently (i.e., poor matching), (2) the performer uses the cues inconsistently, or (3) the listeners use the cues inconsistently. Only by analyzing these three indices separately can one explain the success of the communication in a particular situation (see also Juslin & Scherer, 2005).

Cognitive Feedback

The lens model offers a useful tool for improving communication of emotion in music because it provides explicit knowledge about the relationships among performers, cues, and listeners. This information may be used to provide *cognitive feedback* (CFB). The notion of CFB is to allow a music performer to compare a regression model of his or her playing to an optimal regression model of playing based on listeners' judgments (Juslin & Laukka, 2000).

The term CFB was first introduced in studies of human judgment by Hammond (1971), who provided judges with feedback about task properties and judgment strategies. CFB is usually contrasted with *outcome feedback*, where judges only receive information about whether the judgment was good or bad, but no information about why.

In what way does CFB differ from the kind of feedback commonly provided by music teachers? First, CFB corresponds more closely to the definition of feedback that was given earlier, since it provides a direct comparison of the present strategy with an optimal strategy. Second, CFB differs from teachers' feedback in how the feedback is derived. Many of the performer's manipulations of acoustic cues are audible to listeners in general and to teachers in particular. Yet, it is difficult for a human perceiver to infer the statistical relationships that exist among expressive intentions, acoustic cues, and listener judgments (see Figure 1), let alone the relations among the cues themselves. It is well-known from extensive research on human judgment that judges are commonly unable to explain the basis of their judgments, especially in situations that feature several uncertain cues (Brehmer, 1994). CFB solves this problem by using a statistical method (multiple regression analysis) that makes it possible to describe the complex relationships with a precision that would be hard to achieve for a human perceiver.

A pilot study featuring guitar players at an intermediate level of expertise explored the efficacy of CFB (Juslin & Laukka, 2000). The results showed that CFB yielded about a 50% increase in communication accuracy after a single feedback session, as indicated by listening tests. The regression models of the performers and the listeners in the study were obtained by manually extracting all acoustic cues of the performances and conducting regression analyses using standard software. Such measurements and analyses are complex and time-consuming, wherefore a teaching method that would require teachers to manually extract acoustic cues is not a feasible alternative in music education. Thus, an important goal of the Feel-ME project was to create a computer program that would automatically analyze music performances and provide CFB to performers.

The Feel-ME Program

A first prototype of a computer program that offers CFB to music performers has been developed by the present authors.[1] The current version of the program is implemented using the *Matlab* platform for mathematical computations. The program is organized in terms of four modules, each associated with a distinct user activity in a circular process of recording (*the Recorder*), analyzing acoustic cues (*the Researcher*), receiving feedback (*the Teacher*), and monitoring progress (*Learning Curves*). The goal was to create a program that would be easy to use even for students without much experience of using computers. The design of the program is best illustrated by the user procedure, as described in the following.

In the first phase, the performer is instructed to record several different performances of the same melody in order to communicate various emotions (e.g., happiness, sadness) that are selected at the outset. The performer records several versions of each emotional expression to obtain a representative sample of performances. The performances are stored in the computer memory, and acoustic cues (e.g., tempo, sound level, articulation) are automatically analyzed by the program. Which of these cues are used in further stages of the analysis depends on the particular musical instrument used, since different instruments provide different opportunities for varying each of these cues. Multiple regression analysis is used to model the relationships between the performer's expressive intention and the acoustic cues. This produces indices of consistency (multiple correlation, R_e) and cue weights (correlations or beta weights, β) for the performer. The performer models are also compared to stored regression models of listeners' judgments of emotion in music performances based on previous listening experiments. These listener models are used to simulate new judgments, which are used by the program to obtain indices of achievement and matching. (For details, see Method section.)

In the second phase, the performer requests feedback from the program; this includes a visual and numerical description of the performer's use of cues, the listeners' use of cues, the matching between performer's and listeners' cue weights, the consistency of the performer's use of cues, and the achievement. All this is shown in a graphical interface that resembles the lens model (see Figure 2). This makes it possible to compare directly how performer and listeners use the same cues. Instances of poor matching are highlighted by the black (which appear red in the Juslin et al., 2004, article) arrows that signal that a change in utilization of a cue in a specific direction is recommended. The recommendation is also expressed verbally (e.g., softer). If the performer is using cues in an inconsistent manner, this will be apparent from the consistency index (achievement, matching, and consistency are transformed from correlations to scores from 1 to 5, based on the old Swedish school system). From this point, the performer should try to change his or her use of the cues according to the provided feedback (e.g., to use more legato articulation to communicate sadness).

In the final phase, the performer repeats the first task once again (i.e., recording several music performances that express specific emotions). The program again records and analyzes the acoustic cues of the performances and uses simulated listener judgments to obtain updated lens model indices, which may be compared with previous findings. The aim is to see whether the performer has improved his or her communication by changing the use of cues in the ways recommended by the CFB. By observing the updated CFB, the performer can swiftly examine which cues are used effectively, and which cues need continued attention. This feedback cycle may be repeated as many times as deemed necessary, depending on the goals.

The Feel-ME program has two advantages: First, it is well-suited to the nature of the communicative process, as described by empirical research, because it models the uncertain relationships among intentions, acoustic cues, and judgments, and help to render transparent the communicative process. Indeed, whereas most traditional teaching strategies focus either on acoustic properties (e.g., modeling) or experiential aspects (e.g., metaphors), the Feel-ME program resolves this dualism by describing the relationships between the two. Second, the Feel-ME program comprises the three elements required for deliberate practice: namely (a) a well-defined task, (b) informative feedback, and (c) opportunities for repetition and correction of errors. Although there exist a large number of computer programs for the music profession (for overviews, see Bartle, 1987; Webster, 2002), this is the first program aimed at enhancing communication of emotions.

The Present Study

The purpose of the present study was to evaluate the new computer program. The first aspect of the program that was evaluated was its *performance*—does the program improve a performer's communication? Because software development is a costly and time-consuming endeavor, it was regarded as important to be able to demonstrate modest efficacy, at least, of the program in order to justify the costs of further development. Thirty-six semi-professional jazz/rock guitar players were thus randomly assigned to one of three experimental groups: (1) CFB group, (2) Teacher feedback group, and (3) Contrast group (no feedback). Performance measures were obtained in pre- and posttests. The primary question was how each condition would influence the performers' communication of emotions. Although one could expect all three experimental groups to improve from pre- to posttests (e.g., through practice effects or statistical regression toward the mean), we anticipated significant differences with regard to the degree of improvement.

First, based on the assumption that both the Feel-ME program and music teachers would be able to provide useful feedback to the performers, we predicted that the CFB group and the Teacher group would show a larger improvement in communication accuracy than the Contrast group. Second, assuming that the Feel-ME program would be able to provide more specific feedback to the performers than the music teachers, we predicted that the CFB group would show a larger improvement in communication accuracy than the Teacher group. These predictions were tested on

---

[1] The Feel-ME program was jointly developed by the members of the Feel-ME project. The overall design of the program and the procedure used to obtain CFB were developed by Juslin; the implementation of this design was done by Schoonderwaldt in collaboration with Juslin; the cue extraction algorithm (CUEX) was developed by Friberg in collaboration with Schoonderwaldt and Juslin. Remaining members participated in the testing of the program.
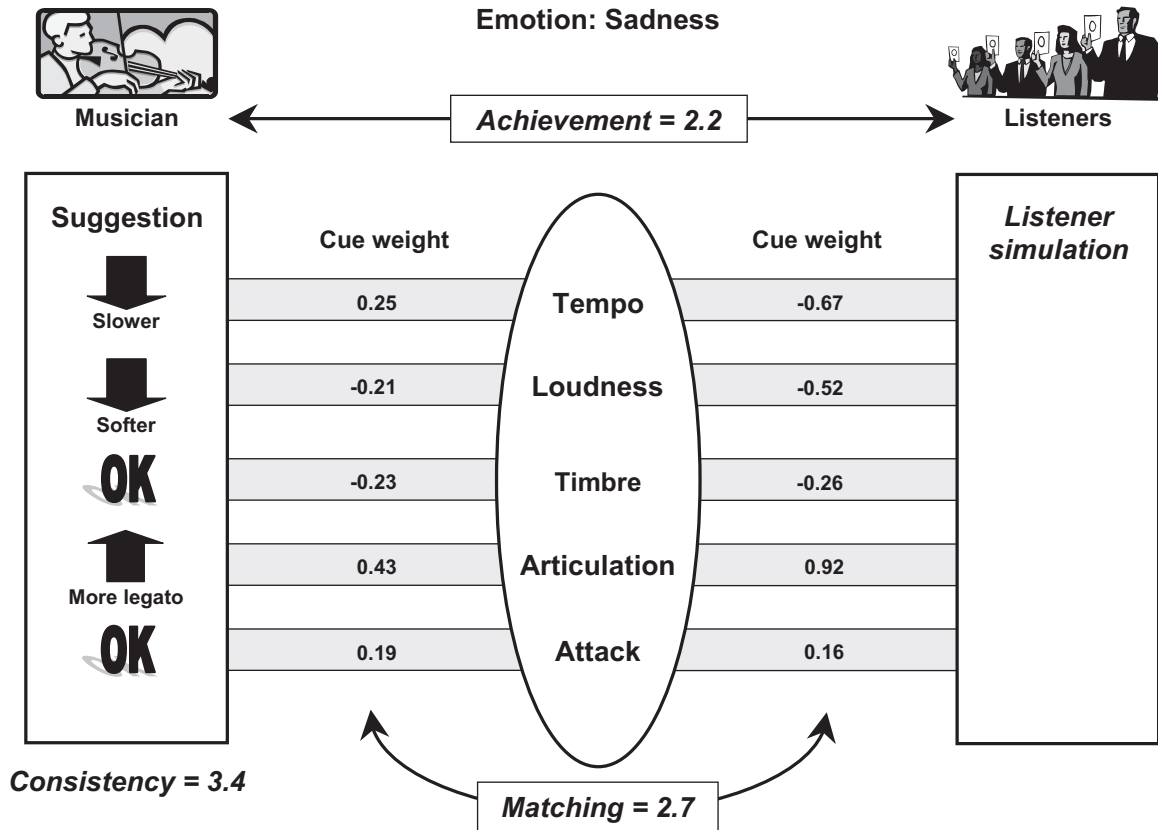
Figure 2. The graphical interface for cognitive feedback featured in the Feel-ME program (from Juslin, Friberg, Schoonderwaldt, & Karlsson, 2004, *Musical excellence* (Feedback learning of musical expressivity), used by permission of Oxford University Press).

performance measures obtained from the Feel-ME program (which was used to record the music performances) and two listening experiments (which allowed us to compare the estimates of the Feel-ME program with listeners' judgments).

The second aspect of the Feel-ME program that was evaluated was its *usability*—is the program user-friendly? It has been recognized that efficacy is not the only important criterion in the evaluation of a novel application. Of equal importance is users' subjective impressions, since if people do not have a favorable reaction to the application, they will not use it anyway (e.g., Balzer et al., 1989). The users' interaction with the program was measured by means of video observation, and their subjective reactions were measured by a questionnaire. Based on previous research on performers' attitudes toward computer-assisted teaching of expressivity (Lindström et al., 2003), we anticipated a negative impression of the program.

Method

Recording Experiment

Performers. Thirty-six semiprofessional guitar players, aged 21 to 49 (*M* = 28), 35 males and one female, participated in the study. They could use their own electric guitar to ensure that they were familiar with the instrument. Their playing experience ranged from 5 to 39 years (*M* = 16.5) and they mainly performed jazz and rock. The performers were paid for

their voluntary and anonymous participation. They were informed that they would be videofilmed during the experiment, and that they could abort the session at any time.

Music teachers. Four guitar teachers, all males, aged 25 to 53 (*M* = 38) participated in the study. They were paid for their anonymous and voluntary participation. The teachers' playing experience ranged from 15 to 40 years (*M* = 24.5). Their teaching experience ranged from 6 to 30 years (*M* = 14.5) and they mainly taught jazz and rock styles. All of them worked professionally as musicians in addition to being teachers at various levels of music education. A questionnaire administered after the experiment showed that all four teachers regarded it as very important to teach expressivity to music students.

Procedure

The performers were randomly assigned to one of three experimental conditions (see below). The basic task was the same in all conditions: The performer was required to play a brief melody, *When the Saints*, so that it would express happiness, sadness, anger, and fear, respectively. These are the four emotions that have been most extensively studied in earlier performance analyses and listening tests (Juslin & Laukka, 2003, Table 3). The results have shown that these emotions are easy to express in music performance, thus ensuring that the emotion labels as such would not prevent reliable communication in this study. However, a different set of labels might be used, since CFB is a general method not tied to any specific emotion label. The piece was chosen because it was short, familiar, easy to play, and highly prototypical of jazz. The performer was asked to play five

versions of each emotion, and to make them as similar as possible. A large number of performances was required in order to obtain a representative sample of performances as well as a reasonable number of cases for subsequent regression analyses. The 20 performances were recorded in both a pretest and a posttest. Each performer thus recorded 40 performances (i.e., 5 versions × 4 emotions × 2 tests). In total, then, 1,440 performances were recorded. The performances were recorded by means of the Feel-ME program, with the guitar connected to a small preamp (*Korg Pandora*) that, in turn, was connected to a computer. The performances were stored as audio files (22 kHz). The recording process was handled by the experimenter, except in the CFB condition where the performer interacted directly with the Feel-ME program.

Communication of emotion involves many cues that are used differently depending on the emotion. To avoid cognitive overload, performers in the feedback groups were instructed to focus on only two of the four emotions during the feedback session. Also, to avoid ceiling effects (e.g., because some participants already had managed to express a particular emotion reliably, thereby making further improvement impossible), feedback sessions focused on the two emotions that the performer had been least successful in expressing initially, as revealed by the achievement in the Feel-ME program; for instance, if a performer in the CFB group or the Teacher feedback group had been least successful in expressing happiness and fear, these were the two emotions that subsequent feedback from program or teacher would focus on. To render results from the three experimental groups comparable, all performance measures in all conditions were taken for the two emotions for which the performer showed the lowest initial achievement. These emotions differed depending on the performer although all emotions were represented in some cases, at least. However, because different emotions were not represented equally, thus rendering comparisons among emotions inappropriate, all performance measures were averaged across emotions in the subsequent data analyses. The remaining features of the procedure were unique to each group, as explained in the following.

Cognitive feedback group. After a brief exploration of the computer program supervised by the experimenter, the performer was required to go through one cycle of CFB, as described previously. The feedback focused on four acoustic cues (i.e., tempo, sound level, articulation, timbre), which have been found to be of crucial importance in communication of emotions in music performance in general (Juslin & Laukka, 2003, Table 7), and electric guitar playing in particular (Juslin, 2000). While the Feel-ME program could include further cues (e.g., vibrato, attack), pilot listening tests revealed that these other cues contributed little predictive power to multiple regression models of listeners' judgments of emotion in electric guitar performances. The performer's interaction with the program was filmed, and the performer also completed a usability questionnaire (described below). The complete experiment took about 2 hours.

Teacher feedback group. The performer carried out the same basic task as in the CFB condition, except that the feedback was now provided by a teacher. There are many teaching strategies that a teacher may use, but research has indicated that verbal instruction dominates in instrumental teaching (e.g., Karlsson & Juslin, 2005; Sang, 1987; Speer, 1994). Therefore, teachers were required to use verbal instruction as much as possible. However, teachers were allowed to use any type of verbal instruction (i.e., metaphors, musical directions, focus on felt emotion, outcome feedback) to help the performer improve his or her communication of each of the two target emotions. Musical modeling (where the teacher plays on an instrument) was not allowed, but physical modeling (e.g., gestures) was allowed since it is a natural part of the verbal instruction process. First, the performer arrived at the laboratory and recorded the first 20 performances. The teacher was not present. After the recording the performer took a break while the experimenter examined which emotions had received the lowest achievement (r_a) in the Feel-ME program. Then, the teacher came to the laboratory and read the instructions. The teacher was asked to listen to the 10 target performances, and to write down verbal feedback that would help

the performer to improve his or her communication of each of the two target emotions. Finally, the performer returned to the laboratory again, where the teacher provided feedback to the performer much as in a regular teaching session. Teacher and performer were videofilmed during the interaction. The teacher instructions were transcribed and coded post hoc by two of the present authors. Intercoder agreement was estimated using Cohen's Kappa (Howell, 1992). Mean intercoder agreement was κ = .92. As can be seen in Table 1, teachers usually combined different types of feedback. The most common type was musical directions, followed by outcome feedback, metaphors, and physical modeling. The complete experiment took about 1½ hours.

Contrast group. The performer received no feedback, but simply performed the musical material twice (pre- and posttest), with a break in between. After the recording, the performer filled out a background questionnaire. The complete experiment took about 1 hour.

Performance Measures

The Feel-ME program computed a number of performance measures that were used to provide CFB to the performers and that also could be used to examine various aspects of the communicative process.

Acoustic measures. Measures of tempo, sound level, articulation, and timbre from the 1,440 performances recorded were automatically analyzed by means of the CUEX algorithm (Friberg, Schoonderwaldt, & Juslin, in press). Each performance is first segmented into tone boundaries through analyses of both sound level and pitch. Potential tone onsets and offsets are detected by finding segments with similar fundamental frequency (pitch) and substantial dips in the sound level. Then, for each detected tone, the following eight acoustic parameters are computed by the algorithm: pitch (in semitones), sound level (dB, upper quartile of sound level within onset-offset), instantaneous tempo (notes per second), articulation (percentage of pause duration), attack velocity (dB/s), spectral balance (dB, difference between high and low spectral content; i.e., a correlate of the perceived timbre), vibrato rate (Hz), and vibrato extent (semitones). The most difficult aspect of the cue extraction is to correctly detect the indi-

Table 1

Post-hoc Categorization of the Music Teachers' Feedback to the Performers

Teacher	Performer	Feedback type				
		Mu	Ou	Me	Mo	Mi
A	1.	10	4	4	6	0
	2.	9	5	3	6	0
	3.	11	3	6	7	0
	Σ_A	30	12	13	19	0
B	4.	6	3	5	0	0
	5.	7	6	4	0	1
	6.	5	4	3	0	2
	Σ_B	18	13	12	0	3
C	7.	6	0	4	0	0
	8.	6	2	8	0	0
	9.	5	1	1	0	0
	Σ_C	17	3	13	0	0
D	10.	5	9	1	1	2
	11.	3	7	4	0	0
	12.	6	6	5	2	0
	Σ_D	14	22	10	3	2
All	Σ_{A-D}	79	50	48	22	4

Note. Mu = musical directions, Ou = outcome feedback, Me = metaphors, Mo = modeling, Mi = miscellaneous (i.e., verbal utterances that could not be easily categorized). Values show the frequency of occurrence for each feedback type.

vidual tones. Preliminary estimates of mean recognition rate, ranging from 90% to 99% (depending on the type of performance sample), reveal that the detection is far from perfect (Friberg et al., in press). However, since subsequent statistical analyses by the Feel-ME program (see below) use only averages of cues across each performance, and rely on correlation statistics, less than perfect note-detection accuracy was not considered a serious problem in this context (see also Friberg, Schoonderwaldt, Juslin, & Bresin, 2002).

Performer models. The acoustic measures were used by the Feel-ME program to create models of each performer's playing strategy. One multiple regression analysis was conducted for each emotional expression by each performer in both pre- and posttest. Thus, no less than 288 (36 performers × 4 emotions × 2 tests) regression analyses were computed. All regression analyses were conducted by means of simultaneous (as opposed to stepwise) linear regression. The performer's expressive intention was the dependent variable and the cues (tempo, sound level, articulation, timbre) were the independent variables; that is, the analyses were designed to reveal how well the intended emotions could be predicted from a linear combination of the cues. The performer's intention was coded dichotomously for each emotion analyzed, so that all performances made with this intention were coded 1, whereas all other performances were coded 0. The cues were coded continuously, using raw data from the acoustic analyses. Each performer model was based on 20 performances (i.e., a case-to-predictor ratio of 5 to 1). The multiple correlation of the regression model was used as a measure of performer consistency. Previous research has indicated that linear regression models provide a good fit to performers' and listeners' utilization of acoustic cues in communication of emotions (Juslin, 1997b; Juslin & Madison, 1999), as could be expected from the lens model (see the introduction).

Cue weights. The Feel-ME program allows a choice of either beta weights or regular correlations as indices of cue weights. In the present study, we chose to use the latter index based on the assumption that it may be easier to interpret for a musician who is not familiar with statistics. Thus, to measure the relations among performers' expressive intentions and cues, the point-biserial correlations (r_{pb}) between the performer's intention and each of the four cues were calculated. The performer's intention was coded dichotomously (see above) and the cues were coded continuously, using the raw data from the acoustic analyses. Thus, for example, the point-biserial correlation between anger intention and mean tempo indexes the extent to which the tempo increases or decreases when the performer intends to express anger (1) as opposed to other emotions (0). This measure was used to index the performer's cue weight for tempo in regard to expression of anger.

Listener Models and Simulation of Judgments

The performer models were related to stored regression models of listeners' judgments by the Feel-ME program. These listener models derived from previous listening experiments in which listeners judged the emotional expressions of a wide range of musical performances with various emotional expressions. (For examples, see Juslin, 1997b, 2000.) Both musically trained and untrained listeners were included, though previous research has indicated that the differences between experts and novices are quite small when it comes to emotion judgments (Juslin, 2001). All models were based on listening tests that featured the same melody as was used in this study to ensure that the models would be suitable for the piece and style. Multiple regression analysis was used to model the relations between listeners' judgments and acoustic cues. The judgments were subjected to one simultaneous regression analysis for each emotion. The mean listener rating on the respective scale was the dependent variable and the cues were the independent variables.

The stored listener models were used to simulate listener judgments through a method called *judgmental bootstrapping* (e.g., Cooksey, 1996). Basically, this means that a multiple regression equation line that was

originally fitted to a sample of cases with certain predictor values is subsequently applied to a new sample of cases with different predictor values. This is slightly similar to a cross-validation procedure in multiple regression. Thus, in the present context, the stored regression model of listeners' emotion judgments was used to predict new listener judgments by entering the cue values from the acoustic analyses (see above) into the existing regression equation. While applying this equation to a new sample may be expected to lead to a drop in predictive accuracy, previous studies suggest that bootstrapping may lead to judgment accuracy equal to or above the accuracy of individual judges (e.g., Dawes, 1982; Dawes & Corrigan, 1974).

Lens model indices. Achievement (r_a) was measured for each emotional expression by each performer in pre- and posttest by the point-biserial correlation between the performer's expressive intention (dichotomously coded) and the predicted listener rating by the Feel-ME program (continuously coded). Matching (G) was measured by the correlation (r) among the predicted values of the performer's regression model and the predicted values of the listeners' regression model. This correlation may be interpreted as the degree to which the performer's cue weights and the listeners' cue weights would agree if both regression models were perfect ($R_e = R_s = 1.0$). Matching is independent of consistency since it is corrected for inconsistency.

Usability

The usability of the Feel-ME program was measured using standard methods from the field of human-computer interaction (Olson & Olson, 2003). The user's interaction with the program was measured by video observation and a questionnaire that also indexed the user's subjective reactions to the program. The questionnaire contained 31 questions. Some questions were inspired by *Questionnaire for User Interface Satisfaction* (Chin, Diehl, & Norman, 1988) and *Nielsen's Attributes of Usability* (Nielsen, 1993). Other questions were particular to the Feel-ME program. The questions addressed aspects such as the naming of commands, the organization of program modules, the consistency of terminology use, as well as more general impressions. Two digital video cameras (*Sony DCR-PC105E*) recorded the performer's interaction with the computer program. The performer sat on a chair in front of the computer. One camera filmed the performer diagonally from the front (angle: 20 degrees; distance: 2.5 m) to record his or her facial expressions and postures. The other camera was placed 1.5 m to the right of the performer, directly facing a second computer screen that projected the same image as the screen in front of the performer, in order to record the performer's navigation in the program. Both video cameras recorded both sound and vision. The performer's screen activity, speech, and behavior were transcribed. First, a rough transcription of the complete session was made. Then, episodes of particular importance (e.g., mistakes) were transcribed in finer detail.

Listening Experiments

Listeners. In Experiment 1, 16 musically trained listeners (university students with experience of playing musical instruments), 9 females and 7 males, 20–34 years old ($M = 28$), participated. In Experiment 2, 14 untrained listeners (university students without any experience of playing musical instruments), 7 females and 7 males, 20–33 years ($M = 25$) participated. The listeners were paid or received course credits for their anonymous and voluntary participation.

Musical Material

The musical material was the same in both experiments and consisted of a subsample of the 1,440 performances recorded. Stimuli were selected in accordance with the procedure used in the feedback sessions (see above). Thus, for each performer, we focused on the two emotions for which the

performer had obtained the lowest initial achievement (according to the computer program). However, because there were as many as five performances of each emotion by each performer in each test (pre/post), some reduction was necessary to obtain a manageable number of stimuli. Hence we randomly selected one performance of each of the two emotions for each performer and test. All together, 144 performances (36 performers × 2 emotions × 2 tests) were included.

Procedure

In Experiment 1, listeners made forced-choice judgments of the performances, which were presented in blocks of pairs with similar intended emotional expressions. Unknown to the listener, one member of each pair was a pretest performance by one of the 36 performers and the other member was a posttest performance by the same performer. The listener's task was simply to judge which of the two versions in each stimulus pair was the most happy (sad, angry, and fearful, respectively). Two randomized stimulus orders were created in which also the order of pre- and postversions within stimulus pairs were randomly distributed across the two stimulus orders. Half the listeners received one stimulus order and the other half received the other stimulus order, according to random assignment.

In Experiment 2, listeners were instructed to rate each stimulus with regard to how well it matched each of the adjectives *happy*, *sad*, *fearful*, and *angry*, on a scale from 0, *not at all*, to 9, *completely*. All emotional expressions were presented in the same block. The order of the stimuli was randomized for each listener. The order in which the adjective scales appeared on the screen was randomized for each listener, but remained the same throughout the session.

Both experiments were conducted individually, using computer software to play sound files (*wav* format) and to record the listeners' responses (pressing buttons or adjusting sliders with the computer mouse). Participants listened to the stimuli through headphones (*AKG-141*) at a comfortable sound level, and they could proceed at their own pace. A break lasting a few minutes was inserted between blocks (Experiment 1), or at stimulus numbers 36, 72, and 108 (Experiment 2). Both experiments took approximately 1.5 hours, including breaks.

Results

Performance Measures

Experiment 1

Figure 3 (upper panel) shows the results from Experiment 1, in which listeners made forced-choice judgments between pre- and posttest versions of performances with the same emotional expression (e.g., which performance sounds more "happy"?). The task of judging which one of two performances best expressed a given emotion was regarded as statistically powerful in detecting even very subtle differences in expression between pre- and posttest in each condition. Moreover, the use of musically trained listeners was expected to increase the sensitivity of the listening experiment. As explained earlier, because different emotions were not equally represented, the results were averaged across emotions. Hence, they indicate the overall extent to which the posttest versions were judged as better or worse exemplars of the intended emotions than were the pretest versions. The primary question in Experiment 1 was how each condition would influence the performers' communication accuracy, as indexed by listeners' judgments. To determine the relative extent of improvement in accuracy among the experimental groups, we performed orthogonal comparisons of the pre/post difference scores. The results are

shown in Table 2 (upper section). Consistent with prediction 1, the CFB group and the Teacher group showed a larger improvement than the Contrast group. This effect was medium ($r_{pb} > .24$), according to Cohen's (1988) guidelines. However, inspection of Figure 3 reveals that the CFB group accounted for most of this effect. This was confirmed by a second comparison, which indicated that, consistent with prediction 2, the CFB group improved more than the Teacher group (see Table 2).

Also shown in Figure 3 is the predicted level of achievement by the Feel-ME program (lower panel). Hence, means of the proportion of pre versus post versions that were selected by listeners in Experiment 1 can be compared with the predicted ratings by the program for the same stimuli. Note that the scales in the upper and lower panels of Figure 3 are different, as they present different types of data. However, the overall patterns can still be compared and are highly similar. Orthogonal comparisons (Table 2, middle section) confirmed that, as was the case for the listener judgments, the Feel-ME program estimated a larger improvement for the CFB group and the Teacher group than for the Contrast group, and a larger improvement for the CFB group than for the Teacher group.

Experiment 2

The findings of Experiment 1 were replicated in Experiment 2, in which other listeners rated each performance on four adjective scales. The judgment task in Experiment 2 arguably provided a less biased estimate of the perceived emotional expression than the judgment task in Experiment 1. First, the intended emotion was not disclosed to the listener. Second, the listener was not forced to choose one emotion. Figure 4 presents listeners' mean ratings of the intended emotion of each performance (across emotions) in pre- and posttest, as a function of condition. We performed orthogonal comparisons of the difference scores (see Table 2, lower section). Again, the CFB group and the Teacher group showed a larger improvement than the Contrast group, and the CFB group showed a larger improvement than the Teacher group.

The effect of CFB was smaller in Experiment 2 than in Experiment 1, perhaps because differences were more difficult to detect in the rating-scale task than in the forced-choice task. Yet, the results suggest that even when all performances with different emotional expressions were presented together in randomized order, and listeners did not know the "right" answer or were forced to choose one performance, they were still able to detect that the performances in the posttest of the CFB condition better conveyed intended emotions than those in the pretest. Thus, the results from Experiments 1 and 2 converge in suggesting that the Feel-ME program was effective in enhancing performers' communication of emotions.

The data from Experiment 2 also made it possible to compare listeners' ratings of each of the 144 music performances on each emotion scale with the Feel-ME program's estimated ratings of these same performances. How well could listeners' actual judgments be predicted based on the computer program's simulated judgments? An overall estimate of the predictive accuracy of the program was obtained by conducting a regression analysis with the listeners' mean ratings of each performance on each scale as the dependent variable, and the program's estimated rating of each performance on each scale as the independent variable. Four emotion scales and 144 performances yielded a total of 576 cases.
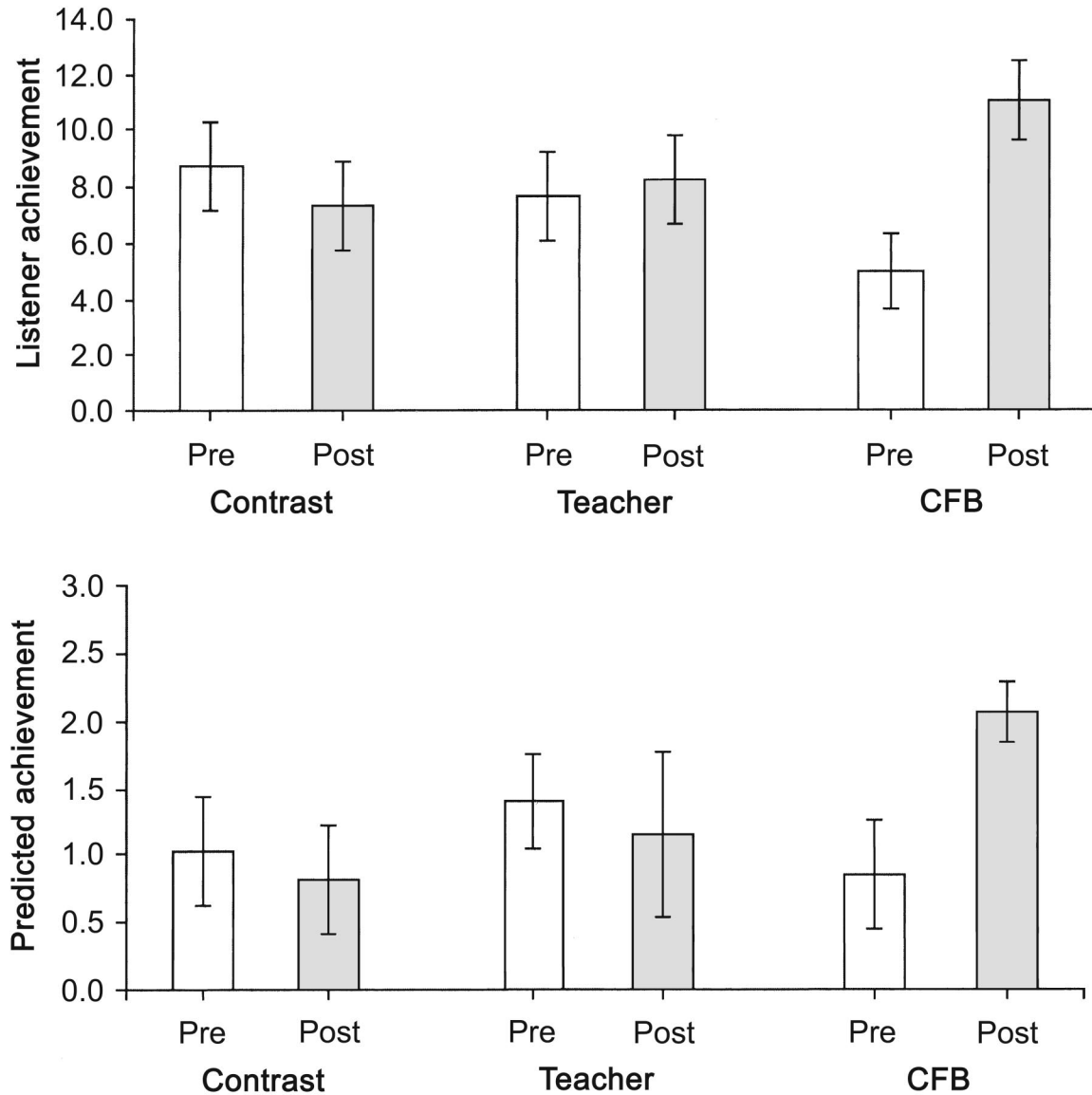
Figure 3. Listeners' forced-choice judgments in Experiment 1 (upper panel), and predicted level of achievement by the Feel-ME program (lower panel), as a function of pre- (light bars) and posttest (dark bars) and experimental condition. Whiskers indicate 95% confidence intervals around the mean.

The regression analysis produced a positive correlation ($R = .61$, $F_{1,559} = 328.76$, $p < .01$, with 15 outliers >2.5 SD removed), but the prediction was far from perfect. It must be noted, however, that a certain loss of predictive accuracy can be expected simply because of the bootstrapping technique (see Method section) that involves applying a multiple regression equation based on one sample of cases to another sample of cases. Considering that multiple regression models of listeners' emotion judgments in previous studies using real music performances have yielded multiple correlations of about $R = .75$ (Juslin, 2000), the R of .61 in the present, bootstrapped (cross-validated) prediction is not surprisingly low. It should also be noted that, whereas the computer program's estimation is based solely on the acoustic properties of the music performances, listeners' judgments are influenced by

other, additional factors which might include guessing based on assumed equal distribution of the emotions implied by the rating scales, effects due to cues not accounted for by the Feel-ME program, as well as fatigue and learning effects during the listening test.

Measures From the Feel-ME Program

The results from Experiments 1 and 2 indicated that the Feel-ME program was effective in enhancing performers' communication of emotions, and further suggested that listeners and program made fairly similar judgments of the performances. Hence, it may be informative to explore in detail the various measures of the communicative process provided by the Feel-ME

Table 2

t Tests of Difference Scores for (a) Listener Judgments in Experiment 1, (b) Achievement Estimated by the Feel-ME Program, and (c) Listener Ratings in Experiment 2

Comparison	M	SD	t	r_{pb}
Listener judgments (Experiment 1)				
CFB/Teacher	3.38	5.16	2.37*	.34
Contrast	−1.48	7.90		
CFB	6.17	6.93	2.69**	.36
Teacher	.58	7.54		
Achievement (Feel-ME program)				
CFB/Teacher	.45	1.12	2.46*	.31
Contrast	−.22	.90		
CFB	1.14	1.71	3.13**	.40
Teacher	−.24	1.39		
Listener ratings (Experiment 2)				
CFB/Teacher	.37	.80	1.89*	.28
Contrast	−.21	1.14		
CFB	.73	1.13	2.54**	.32
Teacher	.01	1.00		

Note. $df = 23$.
* $p < .05$ (one-tailed).   ** $p < .01$ (one-tailed).

program. An advantage of considering the results from the program is that they are based on a larger sample of music performances ($N = 720$) than the listening experiments ($N = 144$).

We conducted orthogonal comparisons of difference scores for achievement, matching, and consistency. Since both analysis and feedback sessions focused on two emotions for each performer (the two emotions for which the performer obtained the lowest initial achievement; see Method section) there were four achieve-

ment scores for each performer (i.e., two from the pretest and two from the posttest). However, because individual scores could not be treated as independent observations, we used mean values across the two emotions for each performer to compute the pre/post difference scores. The results are summarized in Table 3. Beginning with achievement, the results are consistent with those of Experiments 1 and 2 in suggesting that the CFB group and the Teacher group improved more than the Contrast group, and that the CFB group improved more than the Teacher group. The results concerning achievement can be explained by the results concerning matching and consistency (see Table 3). Specifically, the increase in achievement of the CFB group and the Teacher group can be explained by the significant increase in matching from pre- to posttest. In other words, the performers in these groups were able to change their playing in accordance with the listener models. There was a tendency for the Teacher group to show a smaller improvement in consistency than the CFB group, but none of the differences involving consistency reached significance.

Usability Measures

Questionnaire

Tables 4 and 5 present the main results from the usability questionnaire. As can be seen in Table 4, most users reported a favorable impression of the Feel-ME program: they thought it was "rather good" (75%), "rather fun to use" (67%), "very easy to understand" (75%), and "very easy to learn to use" (67%). Of particular importance is that none of the users reported that the program was difficult to understand or learn to use. However, as revealed in Table 5 there was some variability ($SD = 1.80$) in regard to the perceived difficulty of understanding the feedback from the program; 25% of the users experienced that the feedback was difficult to understand (i.e., rating <3). Moreover, 33% of the performers found it difficult to change their playing in accordance
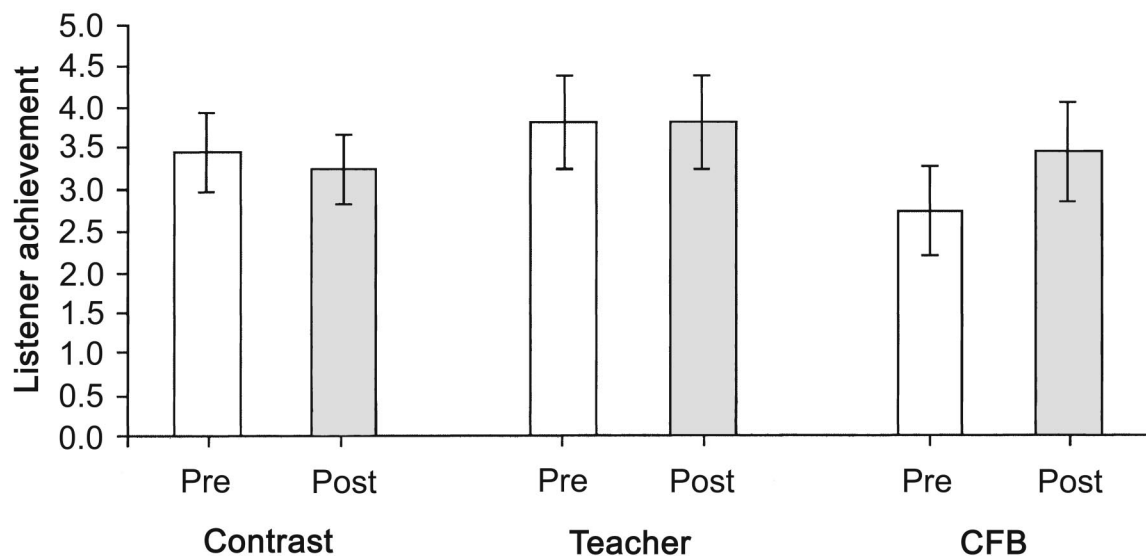

Figure 4. Listeners' mean ratings of intended emotions in Experiment 2 as a function of pre- (light bars) and posttest (dark bars) and experimental condition. Whiskers indicate 95% confidence intervals around the mean.

Table 3
t Tests of Difference Scores for Achievement, Matching, and Consistency Estimated by the Feel-ME Program

Variable	Comparison	M	SD	t	r_{pb}
Achievement (r_a^2)					
	CFB/Teacher	.20	.15	2.93**	.53
	Contrast	.06	.05		
	CFB	.28	.24	2.45*	.39
	Teacher	.12	.11		
Matching (G^2)					
	CFB/Teacher	.21	.20	3.36**	.52
	Contrast	.02	.10		
	CFB	.27	.31	1.35	.23
	Teacher	.15	.17		
Consistency (R_e^2)					
	CFB/Teacher	.11	.14	.40	−.07
	Contrast	.13	.14		
	CFB	.15	.25	.80	.17
	Teacher	.08	.14		

Note. $df = 11$.
* $p < .05$ (one-tailed). ** $p < .01$ (one-tailed).

Table 5
Results From Rating-Scale Questions of the Usability Questionnaire

Question	M	Md	Min	Max	SD
Understanding the feedback suggestions (difficult–easy)	3.60	4.00	2.00	5.00	1.80
Changing playing according to feedback (difficult–easy)	2.80	3.00	1.00	4.00	0.97
Overall grading of the program's quality (low–high)	2.80	3.00	2.00	4.00	0.72
Inclined use of the program in the future (No–Yes)	2.00	1.50	1.00	5.00	1.35
Possibility to improve communication of emotions using the program (No–Yes)	2.30	2.00	1.00	4.00	1.07

Note. Items were rated on a scale from 1 to 5. Anchors are shown within parantheses. $N = 12$.

with the feedback (i.e., rating <3) since it was difficult to change one acoustic parameter without unintentionally changing other parameters also. When asked to rate the overall quality of the program, the modal response was 3 (i.e., neither very low or very high). Further, when asked whether they would consider using the program in the future, 75% of the users responded negatively (i.e., rating <3). This finding may seem surprising in view of the positive impressions that were also reported (see Table 4).

However, the final question shown in Table 5 provides one possible explanation: 58% of the users did not think that the program *can* improve the ability to communicate emotions, and provided comments such as "you cannot learn how to express emotions on an instrument since emotion is a personal thing" and "expression must be honest, it cannot follow a mold". Indeed, reported inclination to use the program in the future was significantly correlated with reported beliefs that the program can improve communication of emotions ($r_{10} = .65$, $p < .05$). Further, users who found it difficult to change their playing strategies in accordance with the provided feedback tended to rate the program more negatively than others ($r_{10} = −.66$, $p < .05$). Still, despite

their skepticism, most users (67%) claimed to have had a high level of ambition in their interactions with the program, which seems indirectly supported by the actual positive outcome with respect to objectives measures of communication accuracy.

Video Observation

Mistakes were categorized as *semantic, syntactic,* or *interactive;* these categories reflect different cognitive levels at which human-computer interaction might occur (Briggs, 1987). A semantic mistake occurs when the user does not understand the logical steps required to solve a particular problem (e.g., not understanding that sound recordings are required in order to get CFB). A syntactic mistake occurs when the user understands the logical steps required, but is unable to map those steps onto the right command facilities available in the program (e.g., not knowing what button to press to start recording). An interactive mistake, finally, occurs when the user knows what to do and how to do it, but simply makes an error in the actual command (e.g., knowing which button to press, but mistakenly pressing another button). Results showed that there were no semantic mistakes, suggesting that the users found the overall design of the Feel-ME program very easy to understand. However, there were 21 syntactic mistakes, which

Table 4
Results From Selected Multiple-Choice Questions of the Usability Questionnaire

Overall impression	n	$\%$	User experience	n	$\%$
Very bad	0	0	Very boring	0	0
Rather bad	3	25	Rather boring	3	25
Rather good	9	75	Rather fun	8	67
Very good	0	0	Very fun	1	8

Understanding the program	n	$\%$	Learning to use the program	n	$\%$
Very difficult	0	0	Very difficult	0	0
Rather difficult	0	0	Rather difficult	0	0
Rather easy	3	25	Rather easy	4	33
Very easy	9	75	Very easy	8	67

Note. $N = 12$.

clearly shows that certain aspects of the design could be improved. Comments in the usability questionnaire and video analyses showed that the syntactic mistakes were primarily due to the misinterpretation of a distinction between *session* (one recording of a set of performances by a performer) and *project* (a minimum of two linked sessions by the same performer). There was merely one interactive mistake. In sum, the video observation confirmed the findings from the usability questionnaire in suggesting that the overall design of the Feel-ME program was easy to understand but that particular aspects of the design could be improved. Such improvements could include a simplified recording procedure, more information about the progress when the program is conducting time-consuming tasks, and a more distinct feedback presentation.

Discussion

In this article, we have presented a novel and empirically based approach to improving communication of emotions in music performance featuring a computer program that records and automatically analyzes musical performances in order to provide feedback to performers. Two listening experiments showed that the program was effective in improving the accuracy of the communicative process. Additional measures from the computer program showed that the improvement in communication accuracy was mainly due to the performers being able to change their performing strategies so that they better matched the optimal models based on listener judgments. Consistent with our first prediction, both the program and feedback from teachers were more effective in improving the communicative process than simple repetition without feedback. Consistent with our second prediction, the results suggested that feedback from the program yielded larger improvements in accuracy than feedback from teachers. One possible explanation of this result is that, whereas the Feel-ME program focused solely on the acoustic cues used to express each emotion, the teachers' feedback often included information that was irrelevant to the task, and that therefore may have been distracting to the performer.

Usability measures showed that the Feel-ME program was favorably perceived by most of the users, but that certain aspects of the design could be improved. It must be noted that the current implementation of the program was done in *Matlab*©, which poses some limits on the graphical design of the program. Thus, the layout of the program could easily be improved in a second prototype. However, there were other problems. Although most users found the program easy to use, some of the less experienced musicians found that the CFB was difficult to understand and use. It has been proposed that the usability of CFB might be affected by the presentation format: graphic, alphanumerical, or verbal; oral or written; immediate or delayed; simple or elaborated (Hammond & Boyle, 1971). However, most studies so far have obtained only minor effects of presentation format (Balzer et al., 1994). A more severe problem in this study was that inexperienced performers found it difficult to separate the individual cues. The Feel-ME program may therefore be most suitable for performers at an intermediate skill level, who are able to manipulate the cues independently, but who have not yet sufficient knowledge about the cue-emotion relationships rendered explicit by the program.

The most striking finding, however, was that most users of the Feel-ME program found it "rather good," "fun to use," "easy to understand," and "easy to learn to use"; yet, when asked whether they would consider using the program if they had the chance, most users responded negatively. This presents us with something of a paradox: the program appears to be working, the users think it is "rather good" and "easy to use," and still they do not want it. However, the comments in the questionnaire suggest that there was a generally negative attitude toward the use of computers to learn expressivity (e.g., what does a computer know about emotions?). If so, this would be consistent with the results from an earlier questionnaire study in which only 20% of the performers surveyed believed that computers might be used to learn expressivity (Lindström et al., 2003).

We argue that the skepticism shown toward computer-assisted teaching of expressivity reflects myths about expression; for instance, that expression cannot be described objectively; that explicit understanding is not beneficial to learning expressivity; and that expressive skills cannot be learned. Hoffren (1964) suggested that music educators by their words attach much importance to expression, but that they are "suspicious of any attempt to study it objectively," claiming it is "too subjective and individualistic for measurement and categorization" (p. 32). It is possible that increased incorporation of theories and findings from research on emotional expression in music performance into the curriculum might lead to a reappraisal.

This study has several important theoretical implications that may contribute to such a reappraisal: First, the study suggests that it is possible to measure objectively the variables that underlie expressive performance. Second, the study demonstrates that contrary to what is sometimes claimed (e.g., Woody, 2000), it is possible to learn expressive skills—provided that one receives informative feedback. Performers are able to make use of explicit feedback concerning individual acoustic cues and to translate such information into altered patterns of playing. Finally, the study suggests that it is possible to de-compose the communication skill into matching and consistency of playing, which both contribute to the accuracy with which a performer conveys emotions to listeners. The findings from this study and a previous study of novices suggest that novices usually need to improve both matching and consistency, whereas experts mainly need to improve matching (Juslin & Laukka, 2000).

Limitations of the Present Study

Although the effectiveness of the Feel-ME program was empirically confirmed by two listening experiments featuring different response formats and different participants, as well as by the performance measures from the program itself, it is clear that the results need to be replicated with other performers, instruments, melodies, methods, and contexts. The efficacy and usability of the program is likely to depend strongly on the individual user, as well as on the specific context of its use. In the present study, musicians were abruptly put in a situation where they had to interact with a computer program in a controlled laboratory setting without prior information about the program's theoretical background; this could perhaps account for some of the attitudes and effects obtained. Thus, a crucial future goal is to test the program in the field in order to increase generalizability in terms of instruments, repertoire, and settings; explore possible long-term benefits; and study individual differences among performers.

There are also several limitations concerning the design of the present experiments. To avoid ceiling effects and cognitive overload, the feedback sessions (and, indeed, most results) focused on the two emotions that each performer was least successful in conveying. Focusing on the lowest and most extreme values of accuracy among the emotions introduces the risk of statistical regression toward the mean, as explained earlier. However, this risk is common to all conditions, and cannot explain why the CFB group improved its communication accuracy more than the other groups, even though the pretest accuracy of all three groups (as shown by two listening experiments and computer program) left room for improvement. A more serious problem resulting from this design is that it precluded all comparisons of the relative efficacy with which the communication of individual emotions could be improved. This issue remains to be investigated, using a more balanced design.

Another limitation of the present study concerns the teacher feedback condition. It may be argued that preventing the teachers from using musical modeling (i.e., imitation of a sound model) rendered the condition unrealistic. It should be noted though that observational studies of instrumental teaching have revealed that little time is devoted to musical modeling during a lesson. Lessons are instead dominated by verbal instruction (Karlsson & Juslin, 2005; see also Sang, 1987; Speer, 1994). Hence, the condition used here may actually be quite representative of what goes on in an instrumental lesson. Even so, it cannot be ruled out that a condition that would have allowed teachers to use musical modeling would have produced different results.

Limitations of the Feel-ME Approach

There are also a number of limitations of the present approach to learning expressivity, more generally. For instance, in its current form the Feel-ME program can only analyze cues from monophonic performances of music (i.e., melody). Thus, the program is mainly suitable for single-line solo instruments such as the violin, flute, guitar, saxophone, and voice; at least until polyphonic extraction of cues is available. The program is also restricted to brief extracts of music. The program analyzes cues only in terms of average measures across each recorded performance and these measures are not meaningful for longer pieces in which the expression may change substantially. One solution to this problem might be to practice different sections of a longer piece in short segments that are suitable for the Feel-ME program. The method of practicing long pieces in short segments is relatively common in instrumental practice (Barry & Hallam, 2002). Another important limitation of the Feel-ME program is that, in the current version at least, the program does not take into account local expressive features that could be important in the expression of emotions (e.g., Juslin & Madison, 1999; Lindström, 2003); nor does it take into account visual features of a performance (e.g., body language, gesture, facial expression) that might convey emotions in a live performance (Ohgushi & Hattori, 1996). The neglect of such features could be one reason that the participants did not feel compelled to use the Feel-ME program in the future.

The Feel-ME program also raises the crucial question of what constitutes an optimal performance. This issue arguably spans many different artistic aspects, including originality, recognition, arousal, beauty, emotion, balance, and personal expression. In the present study, the focus has been on only one of these aspects, namely emotional expression. In the specific context of the Feel-ME program, it is rather easy to define what constitutes an optimal music performance: an optimal performance is one that communicates the intended emotion reliably to listeners by incorporating cues in accordance with how listeners use the same cues in their emotion judgments. Clearly, however, emotional communication should not be the only goal of practice. Performers must develop other aspects as well, using other means (Juslin, 2003). Therefore, an important issue for future research is how different teaching strategies could be effectively combined in more overarching performance interventions (Williamon, 2004).

One final although important limitation of the Feel-ME approach is its dependence on computers. First, not all institutions or individuals have access to computers. Fortunately, recent estimates indicate that the availability of computers in music-educational contexts is increasing (Webster, 2002). Second, computers lack a human touch that may be valued by the student. However, it must be noted that the teacher can play a supporting role also when using computer-assisted teaching strategies, in particular in shaping esthetic judgments and achieving balance among different aspects of expression.

Advantages of the Feel-ME Approach

While acknowledging many potential problems, we also believe there are a number of potential advantages of the present approach in relation to traditional teaching strategies. The Feel-ME program (1) can provide critical feedback but in a nonthreatening environment, (2) is easily available, (3) provides possibilities for flexible and individually based learning, and (4) explicitly describes relationships among expressive intentions, acoustic cues, and listener impressions that are typically embedded in tacit knowledge. The time required to go through one cycle of CFB (as outlined in the Introduction) is approximately the same as that required by a regular music lesson (i.e., 40–60 minutes).

The Feel-ME approach offers a certain level of generality since the basic procedure of CFB (recording, analysis, simulation of listener judgments, feedback) could, in principle, be used with any style of music. What is needed to adapt the program to a particular style is (1) that all acoustic cues that are relevant to the style are included in the analysis and (2) that the regression models used to predict listeners' judgments are based on listening experiments in which musical examples, emotion terms, and listeners are appropriate for the musical genre. Although one could fear that use of the Feel-ME program could lead to a standardization of performances of music, it must be noted that the decision about how to interpret the music is left to the performer. The Feel-ME program only serves to help performers achieve intended musical interpretations more reliably, whatever those may be, by giving performers a deeper understanding of the relationships among expressive features and perceptual effects.

Besides being a potentially useful practice tool, the Feel-ME program could also serve as a diagnostic test of expressive skills for musicians and music teachers (cf. Hoffren, 1964). There is some evidence showing that inexperienced music teachers are less able to diagnose performance problems concerning emotional expression than are expert teachers (Doerksen, 1999). The Feel-ME program could assist teachers in identifying weaknesses with re-

spect to specific aspects of a performer's expressive strategy. Because the Feel-ME program provides many indices of the communicative process, it could also be used to study learning processes in emotional expression in music performance. Finally, the Feel-ME program could become a valuable research tool, because it can help music researchers to swiftly analyze the expressive features of large samples of music performances (Friberg et al., in press).

Concluding Remarks

To conclude, the present study has suggested that it is possible to construct a computer program that automatically analyzes the acoustic cues of music performances, creates models of playing strategies, and provides informative feedback to performers that can improve their communication of emotions. It is only quite recently that a computer program of this type has become possible, thanks to (a) increased formal knowledge about communication of emotions in music performance and (b) unprecedented levels of processing speed in personal computers required for the complicated computations. Both the present study and other studies that have compared computer-assisted teaching with traditional teaching suggest that computer-assisted teaching can be effective (Webster, 2002). Whatever the limitations of the Feel-ME method or this study, the results clearly indicate that computer-assisted teaching of emotional expression is a promising avenue that is worth further development and evaluation. Such evaluation will have to address the crucial question, left unanswered by this study, whether the benefits of the new music technology will exceed the costs.

References

Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin, 106,* 410–433.

Balzer, W. K., Hammer, L. B., Sumner, K. E., Birchenough, T. R., Parham Martens, S., & Raymark, P. H. (1994). Effects of cognitive feedback components, display format, and elaboration on performance. *Organizational Behavior and Human Decision Processes, 58,* 369–385.

Barry, N. H., & Hallam, S. (2002). Practice. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance: Creative strategies for teaching and learning* (pp. 151–165). New York: Oxford University Press.

Barten, S. S. (1998). Speaking of music: The use of motor-affective metaphors in music instruction. *Journal of Aesthetic Education, 32,* 89–97.

Bartle, B. K. (1987). *Computer software in music and music education: A guide.* Metuchen, NJ: Scarecrow.

Boyd, J., & George-Warren, H. (1992). *Musicians in tune: Seventy-five contemporary musicians discuss the creative process.* New York: Fireside.

Brehmer, B. (1994). The psychology of linear judgement models. *Acta Psychologica, 87,* 137–154.

Briggs, P. (1987). Usability assessment for the office: Methodological choices and their implications. In M. Frese, E. Ulich, & W. Dizda (Eds.), *Human computer interaction in the workplace* (pp. 381–401). Amsterdam: Elsevier.

Brunswik, E. (1956). *Perception and the representative design of experiments.* Berkeley, CA: University of California Press.

Bruser, M. (1997). *The art of practicing: A guide to making music from the heart.* New York: Bell Tower.

Budd, M. (1985). *Music and the emotions: The philosophical theories.* London: Routledge.

Budd, M. (1989). Music and the communication of emotion. *Journal of Aesthetics and Art Criticism, 47,* 129–138.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988, May). Development of an instrument for measuring user satisfaction of the human-computer interface. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 213–218). New York: ACM Press.

Clarke, E. F. (1988). Generative principles in music performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 1–26). Oxford, England: Clarendon Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Cooksey, R. W. (1996). *Judgment analysis: Theory, method, and applications.* New York: Academic Press.

Dalgarno, G. (1997). Creating an expressive performance without being able to play a musical instrument. *British Journal of Music Education, 14,* 163–171.

Davies, S. (1994). *Musical meaning and expression.* Ithaca, NY: Cornell University Press.

Dawes, R. M. (1982). The robust beauty of improper linear models in decision making. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 391–407). Cambridge, England: Cambridge University Press.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81,* 95–106.

Dickey, M. R. (1992). A review of research on modeling in music teaching and learning. *Bulletin of the Council for Research in Music Education, 113,* 27–40.

Doerksen, P. F. (1999). Aural-diagnostic and prescriptive skills of preservice and expert instrumental music teachers. *Journal of Research in Music Education, 47,* 78–88.

Ebie, B. D. (2004). The effects of verbal, vocally modeled, kinesthetic, and audio-visual treatment conditions on male and female middle-school vocal music students' abilities to expressively sing melodies. *Psychology of Music, 32,* 405–417.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406.

Friberg, A., Schoonderwaldt, E., & Juslin, P. N. (in press). CUEX: An algorithm for extracting expressive tone parameters in music performance. *Acta Acustica united with Acustica.*

Friberg, A., Schoonderwaldt, E., Juslin, P. N., & Bresin, R. (2002, September). Automatic real-time extraction of musical expression. In *Proceedings of the International Computer Music Conference, Göteborg* (pp. 365–367). San Francisco: International Computer Music Association.

Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 503–534). New York: Oxford University Press.

Gellrich, M. (1991). Concentration and tension. *British Journal of Music Education, 8,* 167–179.

Hammond, K. R. (1971). Computer graphics as an aid to learning. *Science, 172,* 903–908.

Hammond, K. R., & Boyle, J. R. (1971). Quasi-rationality, quarrels, and new conceptions of feedback. *Bulletin of the British Psychological Society, 24,* 103–113.

Hepler, L. E. (1986). *The measurement of teacher/student interaction in private music lessons and its relation to teacher field dependence/field independence.* Unpublished doctoral dissertation, Case Western Reserve University.

Hoffren, J. (1964). A test of musical expression. *Council for Research in Music Education, 2,* 32–35.

Howell, D. C. (1992). *Statistical methods for psychology* (3rd ed.). Belmont, CA: Duxbury Press.

Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological Review, 71,* 42–60.

Johnson, C. M. (1998). Effect of instruction in appropriate rubato usage on the onset timings and perceived musicianship of musical performances. *Journal of Research in Music Education, 46,* 436–445.

Juslin, P. N. (1995). Emotional communication in music viewed through a Brunswikian lens. In G. Kleinen (Ed.), *Music and expression: Proceedings of the Conference of DGM and ESCOM, Bremen* (pp. 21–25). Bremen, Germany: University of Bremen.

Juslin, P. N. (1997a). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception, 14,* 383–418.

Juslin, P. N. (1997b). Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy. *Musicae Scientiae, 1,* 225–256.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1797–1813.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–337). New York: Oxford University Press.

Juslin, P. N. (2003). Five facets of musical expression: A psychologist's perspective on music performance. *Psychology of Music, 31,* 273–302.

Juslin, P. N. (2005). From mimesis to catharsis: Expression, perception, and induction of emotion in music. In D. Miell, R. MacDonald, & D. J. Hargreaves (Eds.), *Musical communication* (85–115). New York: Oxford University Press.

Juslin, P. N., Friberg, A., & Bresin, R. (2002). Toward a computational model of expression in music performance: The GERM model. *Musicae Scientiae, Special Issue 2001–2002,* 63–122.

Juslin, P. N., Friberg, A., Schoonderwaldt, E., & Karlsson, J. (2004). Feedback-learning of musical expressivity. In A. Williamon (Ed.), *Musical excellence: Strategies and techniques for enhancing performance* (pp. 247–270). New York: Oxford University Press.

Juslin, P. N., & Laukka, P. (2000). Improving emotional communication in music performance through cognitive feedback. *Musicae Scientiae, 4,* 151–183.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research, 33,* 217–238.

Juslin, P. N., Laukka, P., Friberg, A., Bresin, R., & Lindström, E. (2001, August). *Play it again with feeling: Feedback learning of musical expression.* Poster presented at Feelings and Emotions: The Amsterdam Symposium, Amsterdam, the Netherlands.

Juslin, P. N., & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception, 17,* 197–221.

Juslin, P. N., & Persson, R. S. (2002). Emotional communication. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance: Creative strategies for teaching and learning* (pp. 219–236). New York: Oxford University Press.

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 65–135). New York: Oxford University Press.

Karlsson, J., & Juslin, P. N. (2005, September). *The problem of musical expression: An observational study of instrumental teaching.* Paper presented at Performance Matters: International Conference on Psychological, Philosophical, and Educational Issues in Musical Performance, Porto.

Laukka, P. (2004). Instrumental music teachers' views on expressivity: Preliminary evidence from music conservatoires. *Music Education Research, 6,* 45–56.

Lehmann, A. C. (1997). Acquired mental representations in music performance: Anecdotal and preliminary empirical evidence. In H. Jørgensen & A. C. Lehmann (Eds.), *Does practice make perfect? Current theory and research on instrumental music practice* (pp. 141–163). Oslo: Norwegian State Academy of Music.

Lindström, E. (2003). The contribution of immanent and performed accents on emotional expression in short tone sequences. *Journal of New Music Research, 32,* 269–280.

Lindström, E., Juslin, P. N., Bresin, R., & Williamon, A. (2003). "Expressivity comes from within your soul": A questionnaire study of music students' perspectives on expressivity. *Research Studies in Music Education, 20,* 23–47.

Marchand, D. J. (1975). A study of two approaches to developing expressive performance. *Journal of Research in Music Education, 23,* 14–22.

Matthay, T. (1913). *Musical interpretation: Its laws and principles, and their application in teaching and performing.* Boston: Boston Music.

Minassian, C., Gayford, C., & Sloboda, J. A. (2003, March). *Optimal experience in musical performance: A survey of young musicians.* Paper presented at the Meeting of the Society for Education, Music, and Psychology Research, London.

Nielsen, J. (1993). *Usability engineering.* Boston: Academic Press.

Ohgushi, K., & Hattori, M. (1996). Emotional communication in performance of vocal music. In B. Pennycook & E. Costa-Giomi (Eds.), *Proceedings of the Fourth International Conference on Music Perception and Cognition* (pp. 269–274). Montreal, Canada: ICMPC.

Olson, G. M., & Olson, J. S. (2003). Human-computer interaction: Psychological aspects of the human use of computing. *Annual Review of Psychology, 54,* 491–516.

Persson, R. S. (1993). *The subjectivity of musical performance: An exploratory music-psychological real world enquiry into the determinants and education of musical reality.* Doctoral dissertation, University of Huddersfield, Huddersfield, England.

Persson, R. S. (1996). Concert musicians as teachers: On good intentions falling short. In A. J. Cropley & D. Dehn (Eds.), *Fostering the growth of high ability: European perspectives* (pp. 303–320). Norwood, NJ: Ablex.

Ratner, L. G. (1980). *Classic music: Expression, form, and style.* New York: Schirmer.

Reimer, B. (2003). *A philosophy of music education* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Rosenberg, H. S., & Trusheim, W. (1989). Creative transformations: How visual artists, musicians, and dancers use mental imagery in their work. In P. Robin & J. E. Schorr (Eds.), *Imagery: Current perspectives* (pp. 55–75). New York: Plenum Press.

Rostvall, A.-L., & West, T. (2001). *Interaktion och kunskapsutveckling* [Interaction and learning. A study of music instrument teaching]. Unpublished doctoral dissertation, KMH Förlaget, Stockholm.

Sang, R. C. (1987). A study of the relationship between instrumental music teachers' modeling skills and pupil performance behaviors. *Bulletin of the Council for Research in Music Education, 91,* 155–159.

Seashore, C. E. (1938). *The psychology of music.* New York: McGraw-Hill.

Serafine, M. L. (1980). Against music as communication: Implications for music education. *Journal of Aesthetic Education, 14,* 85–96.

Shove, P., & Repp, B. H. (1995). Musical motion and performance:

Theoretical and empirical perspectives. In J. Rink (Ed.), *The practice of performance: Studies in musical interpretation* (pp. 55–83). New York: Cambridge University Press.

Sloboda, J. A., Minassian, C., & Gayford, C. (2003). *Assisting advanced musicians to enhance their expressivity: An intervention study.* Paper presented at the Fifth Triennial Conference of the European Society for the Cognitive Sciences of Music, Hanover, Germany.

Speer, D. R. (1994). An analysis of sequential patterns of instruction in piano lessons. *Journal of Research in Music Education, 42,* 14–26.

Tait, M. (1992). Teaching strategies and styles. In R. Cowell (Ed.), *Handbook of research on music teaching and learning* (pp. 525–534). New York: Schirmer.

Thompson, W. F., & Robitaille, B. (1992). Can composers express emotions through music? *Empirical Studies of the Arts, 10,* 79–89.

Webster, P. R. (2002). Computer-based technology and music teaching and learning. In R. Colwell & C. Richardson (Eds.), *The new handbook of research on music teaching and learning* (pp. 416–439). New York: Oxford University Press.

Williamon, A. (2004). Epilogue. In A. Williamon (Ed.), *Musical excellence: Strategies and techniques to enhance performance* (pp. 291–292). Oxford, England: Oxford University Press.

Woody, R. H. (1999). The relationship between explicit planning and expressive performance of dynamic variations in an aural modeling task. *Journal of Research in Music Education, 47,* 331–342.

Woody, R. H. (2000). Learning expressivity in music: An exploratory study. *Research Studies in Music Education, 14,* 14–23.