Visual Perception of Expressiveness in Musicians' Body Movements

Sofia Dahl and Anders Friberg
KTH, Royal Institute of Technology
Stockholm, Sweden

Musicians often make gestures and move their bodies expressing a musical intention. In order to explore to what extent emotional intentions can be conveyed through musicians' movements, participants watched and rated silent video clips of musicians performing the emotional intentions Happy, Sad, Angry, and Fearful. In the first experiment participants rated emotional expression and movement character of marimba performances. The results showed that the intentions Happiness, Sadness, and Anger were well communicated, whereas Fear was not. Showing selected parts of the player only slightly influenced the identification of the intended emotion. In the second experiment participants rated the same emotional intentions and movement character for performances on bassoon and soprano saxophone. The ratings from the second experiment confirmed that Fear was not communicated whereas Happiness, Sadness, and Anger were recognized. The rated movement cues were similar in the two experiments and were analogous to their audio counterpart in music performance.

Received February 26, 2003, accepted December 4, 2006.

Key words: expression, movement perception, music performance, emotions, gesture analysis

Body movement is an important non-verbal means of communication between humans. Body movements can help observers extract information about the course of action, or the intent of a person. Some of this information is very robust and can be perceived even when certain parts of the moving body are occluded. Such information can even be perceived if the movement is shown just by single light points fastened to the body and displayed with high contrast to give a discrete-point impression (*point-light* technique, see Johansson, 1973). It has been shown that by viewing motion patterns, participants are able to extract a number of non-trivial features such as the sex of a person, the weight of the box he/she is carrying (Runeson & Frykholm, 1981), or landing positions of strokes of badminton playing (Abernethy & Russel, 1987). It is also possible to identify the emotional expression in dance and music performances (Dittrich, Troscianko, Lea, & Morgan, 1996; MacFarlane, Kulka, & Pollick, 2004; Sörgjerd, 2000; Walk & Homan, 1984), as well as the emotional expression in everyday arm movements such as drinking and lifting (Paterson, Pollick, & Sanford, 2001; Pollick, Paterson, Bruderlin, & Sanford, 2001).

Music has an intimate relationship with movement in several aspects. The most obvious relation is that all sounds from traditional acoustic instruments are produced by human movement. Some characteristics of this motion will inevitably be reflected in the resulting tones. For example, the sound level, amplitude envelope, and spectrum change produced on a violin has a direct relationship with the velocity and pressure during the bow gesture (e.g., Askenfelt, 1989). Also, the striking velocity in drumming is strongly related to the height to which the drumstick is lifted in preparation for the stroke (Dahl, 2000, 2004).

Musicians also move their bodies in a way that is not directly related to the production of tones. Head shakes or body sway are examples of movements that, although not actually producing sound, still can serve a communicative purpose of their own. In studies of speech production, McNeill et al. (2002) have argued that speech and movement gestures arise from a shared semantic source. In this respect the movements and the spoken words are co-expressive, one is not subordinate to the other. Bearing in mind that music also is a form of communication and that speech and music have many properties in common (see e.g., Juslin & Laukka, 2003), it is plausible that a similar concept applies to musical communication as well. In earlier studies of music performance the body gestures not directly involved in the production of notes have been referred to as *ancillary*, *accompanist*, or *non-obvious* movements (e.g., Wanderley, 2002). We prefer to think of these performer movements as a *body language* since, as we will see below, they serve several important functions in music performance. It seems reasonable to assume that

Music Perception volume 24, issue 5, pp. 433–454, issn 0730-7829, electronic issn 1533-8312 DOI:10.1525/mp.2007.24.5.433

some of the expressivity in the music is reflected in these movements.

The body movements may also be used for more explicit communication. Davidson and Correia (2002) suggest four aspects that influence the body language in musical performances: (1) Communication with co-performers, (2) individual interpretations of the narrative or expressive/emotional elements of the music, (3) the performer's own experiences and behaviors, and (4) the aim to interact with and entertain an audience. Separating the influence of each of these aspects on a specific movement may not be possible in general. In the present study we have chosen to focus on the individual interpretations and behaviors. By concentrating on solo performances without an audience, aspects (2) and (3) may be dominating and the influences of aspects (1) and (4) would be minimized.

It is well documented that a viewer can perceive expressive nuances from a musician's body language only. Davidson has conducted several studies on expressive movements in musical performance relating the overall perceived expressiveness to musicians' movements (e.g., Clarke & Davidson, 1998; Davidson, 1993, 1994). Most of these studies used video recordings, utilizing the point-light technique (Johansson, 1973) to capture the movements of violinists or pianists. They were instructed to play with three different expressive intentions: deadpan, projected, and exaggerated; instructions that were assumed to be commonly used in music teaching. Participants rated these performances on a scale of expressiveness (ranging from "inexpressive" to "highly expressive"). Davidson (1993) concluded that participants were about equally successful in identifying the expressive intent regardless of whether they were allowed to only listen, only watch, or both watch and listen. Musically naïve participants even performed better when only watching, compared to the other conditions (Davidson, 1995). Similar results were reported by Shinosako and Ohgushi (1996), thus implying that many listeners at a concert may grasp the expressiveness of the performance mainly from the artist's gestures rather than from the musical content.

Davidson (1994) also investigated which parts of a pianist's body conveyed the information used for judging expressiveness. Using the same point-light technique as in other studies, presenting single or different combinations of the points, she found that the head was important for the observers to discriminate between deadpan, projected, or expressive performances, whereas the hands were not.

Sörgjerd (2000) found that the player's intended emotional expression was reflected in the body motion and could be decoded by participants. One clarinet player and one violinist performed pieces with the emotional intentions Happiness, Sadness, Anger, Fear, Solemnity, Tenderness, and No Expression. Participants were asked to select the most appropriate emotion for each performance. Sörgjerd found that participants were better in identifying the emotions Happiness, Sadness, Anger, and Fear than Tenderness and Solemnity. There were no significant differences between the presentation conditions watch-only, listen-only, or both-watch-and-listen. In the watch-only condition, the correct emotion was more often identified for the violinist than for the clarinetist.

In view of the reported ability to discriminate between different expressive intentions, an interesting question to ask is what makes this discrimination possible. What types of movements supply the bits of information about the intent and mood of a performer? Which *movement cues* are used?

Boone and Cunningham (2001) found that children as young as 4 and 5-years old used differentiated movement cues when asked to move a teddy bear to Angry, Sad, Happy, and Fearful music. For the Sad music the children used less force, less rotation, slower movements, and made fewer shifts in movement patterns than they used for the other emotions. The children also used more upward movement for the Happy and Angry music than for Fearful (which, in turn, received more upward movement than Sad music). The accuracy of children's ability to communicate the emotional content to adult observers was strongest for Sad and Happy music and weaker for Angry and Fearful music.

De Meijer (1989, 1991) and Boone and Cunningham (1999) proposed several movement cues considered important for detecting emotional expression (see overview in Boone & Cunningham, 1998). These cues include frequency of upward arm movement, the amount of time the arms were kept close to the body, the amount of muscle tension, the amount of time an individual leaned forward, the number of directional changes in face and torso, and the number of tempo changes an individual made in a given action sequence. The proposed cues are well matched to the findings by De Meijer (1989, 1991), concerning viewers' attribution of emotion to specific body movements. For instance, he found that observers associated actors' performances with Joy if the actors' movements were fast, upward directed, and with arms raised. Similarly the optimal movements for Grief were slow, light, downward directed, and with arms close to the body.

Similarly, Camurri, Lagerlöf, and Volpe (2003) found a connection between the intended expression of dance and the extent to which the limbs are kept close to the body. In their study, automatic movement detection was used to extract cues in rated dance performances with the expressive intentions Joy, Anger, Fear, and Grief. The cues studied were amount of movement (quantity of motion), and how contracted the body was; that is, how close the arms and legs are to the center of gravity (contraction index). They found that performances of Joy were fluent with few movement pauses and with the limbs outstretched. Fear, in contrast, had a high contraction index, i.e., the limbs were often close to the center of gravity.

That the direction of movement and the arm movements seem to be of such importance for perceiving expression in dance is interesting in light of the previously mentioned studies using musicians' movements. The arm movements of a musician are primarily for sound production and thus expressive body language cannot be allowed to interfere if the performance is to be musically acceptable. Thus the expressive movement cues used by the observers to detect emotional expression must either appear in other parts of the body, or *coincide* with the actual playing movements.

The studies mentioned above have all brought up different aspects of the visual link between performer and observer. An interesting comparison can be made with how musical expressiveness is encoded and decoded in the sound. In analysis of music performances, Gabrielsson and Juslin (Gabrielsson & Juslin, 1996; Juslin, 2000, 2001) have explored what happens when a musician performs the same piece of music with different emotional intentions. A set of acoustical cues has been identified (such as tempo, sound level, etc.) that listeners utilize when discriminating between different performances. For example, a Happy performance is characterized by fast mean tempo, high sound level, staccato articulation, and fast tone attacks, whereas a Sad performance is characterized by slow tempo, low sound level, legato articulation, and slow tone attacks. It seems reasonable to assume that the body movements in the performances contain cues corresponding to those appearing in the audio signal. After all, the movements are intimately connected to the sound production. Many of the cues used to characterize music performances intuitively have a direct motional counterpart if we assume that a tone corresponds to a physical gesture: tempo - gesture rate, sound level - gesture size, staccato articulation - fast gestures with a resting part, tone attack-initial gesture speed.

Another coupling between motion and music is that music listening may evoke an imaginary sense of motion (e.g., Clarke, 2001; Shove & Repp, 1995). Similar to visual illusion or animation, changes in pitch, timbre, and dynamics in music would have the capacity of specifying movement. Many factors in music performance have been suggested to influence and evoke this sense of motion. Rhythmic features are a natural choice, as indicated by performance instructions such as *andante* (walking), or *corrente* (running). Some experimental data point in this direction. Friberg and Sundberg (1999) found striking similarities between velocity curves of stopping runners and the tempo curves in final *ritardandi*. Similarly, Juslin, Friberg, and Bresin (2002) found that synthesized performances obtained significantly higher ratings for the adjectives Gestural, Human, Musical, and Expressive when the phrases had a tempo curve corresponding to a model of hand gesture velocity.

Why and when are we experiencing motion in music listening? From a survival point-of-view, Clarke (2001) argues that all series of sound events may evoke a motion sensation since we are trained to recognize physical objects in our environment and deduce the motion of these objects from the sound. Considering the indefinite space of different sounds and sound sequences emanating from real objects, it is plausible that we make a perceptual effort to translate all sound sequences to motion. Todd (1999) even suggests that the auditory system is directly interacting with the motor system in such a way that an imaginary movement is created directly in motor centra. Since performers are listening to their own performances this implies that there is a loop between production and perception and that the body expression must have a close connection with the music expression.

In this study, the main objective was to find out if expressive communication of specific emotions in music performance is possible using body movements only (i.e., excluding the auditory information). A second objective was to find out whether this communication can be described in terms of movement cues (such as slow - fast, jerky - smooth, etc.), similar to those appearing when listening to music performances. A number of different aspects of musicians' body movements have been identified above. We assume in this investigation that the body movement of the player mainly consists of movements for the direct sound production on the instrument, and natural expressive movements not primarily intended to convey visual information to the audience or to fellow musicians.

The specific questions addressed were the following:

1. How successful is the overall communication of each intended emotion?
2. Are there any differences in the communication depending on what part of the player the observers see?
3. How can perceived emotions be described in terms of movement cues?
4. Are there differences depending on performer?

Two experiments were performed to answer these questions. In Experiment 1 participants rated marimba performances, and in Experiment 2 participants rated woodwind performances.

Experiment 1

In the first experiment a percussionist performed a short piece with differing emotional intentions. Based on the assumption that seeing less of the performer would affect the communication, and that some parts of the player would be more important to convey the intention than others, the participants were presented with video clips showing the player to different extent. Preliminary results of Experiment 1 were reported in Dahl and Friberg (2004).

Method

STIMULUS MATERIAL

A professional percussionist was asked to prepare performances of a piece for marimba with four different expressive intentions: Anger, Happiness, Sadness, and Fear. She was instructed to perform the different emotions in a natural, musical way. Thus, implicitly the instructions clearly concerned auditory, not physical, expression. The player was aware that the performances would be video recorded, but not how they were going to be analyzed. No instructions concerning movements or performance manner were given.

The piece chosen was a practice piece from a study book by Morris Goldenberg: “Melodic study in sixteens”(Goldenberg, 1950). This piece was found to be of a suitable duration and of rather neutral emotional character, allowing different interpretations. The player estimated that a total of 5 hours was spent in the preparation for the performance and for the recording.

The recording was carried out using a digital video camera (SONY DCR-VX1000E) placed on a stand at a fixed distance in front of the player. No additional lightning was used in the room (a practice studio at the Royal College of Music, Stockholm), and the camera’s automatic settings were used.

The experimenter checked that the player was clearly in view and prepared the camera for recording, but was not present in the room during the recording. The player performed each intention twice with a short pause between each performance. Afterwards, the player reported that she prepared for the next performance during these pauses by recalling memories of situations where she had experienced the intended emotion. Informal inspection of the video material by the authors and other music researchers suggested that the music expressed the intended emotions and that the body was moving in a natural, not exaggerated way.

The original video files were edited using VirtualDub, a freeware video editing software (Lee, 2006). To remove facial expressions a threshold filter was used, transforming the color image to a strict black and white image (without gray scales). Different *viewing conditions* were prepared, showing the player to a varying degree. Four viewing conditions were used; *full* (showing the full image), *nohands* (the player’s hands not visible), *torso* (player’s hands and head not visible), and *head* (only the player’s head visible). The four conditions were cut out from the original full scale image, using a cropping filter. Figure 1 shows the four viewing conditions for one frame. Based on the original eight video recordings a total of 32 (4 Emotions $\times$ 2 Repetitions $\times$ 4 Conditions) video clips were generated. The duration of the video clips varied between 30 and 50 s.

PARTICIPANTS

A total of 20 participants (10 male and 10 female), mostly students and researchers at the department, volunteered to participate in the experiment. The participants did not receive any economical compensation for their participation. The participants were between 15 and 59 years old (M = 34, SD = 13.6) with varying amounts of music training. Seven participants reported that they had never played a musical instrument, seven participants had played a musical instrument previously, and six participants had experience of playing one or many musical instruments for many years and currently played between 1 and 6 hours per week.

PROCEDURE

Participants were asked to rate the emotional content in the video clips on a scale from 0 (“nothing”) to 6 (“very much”), for the four emotions Fear, Anger, Happiness, and Sadness.

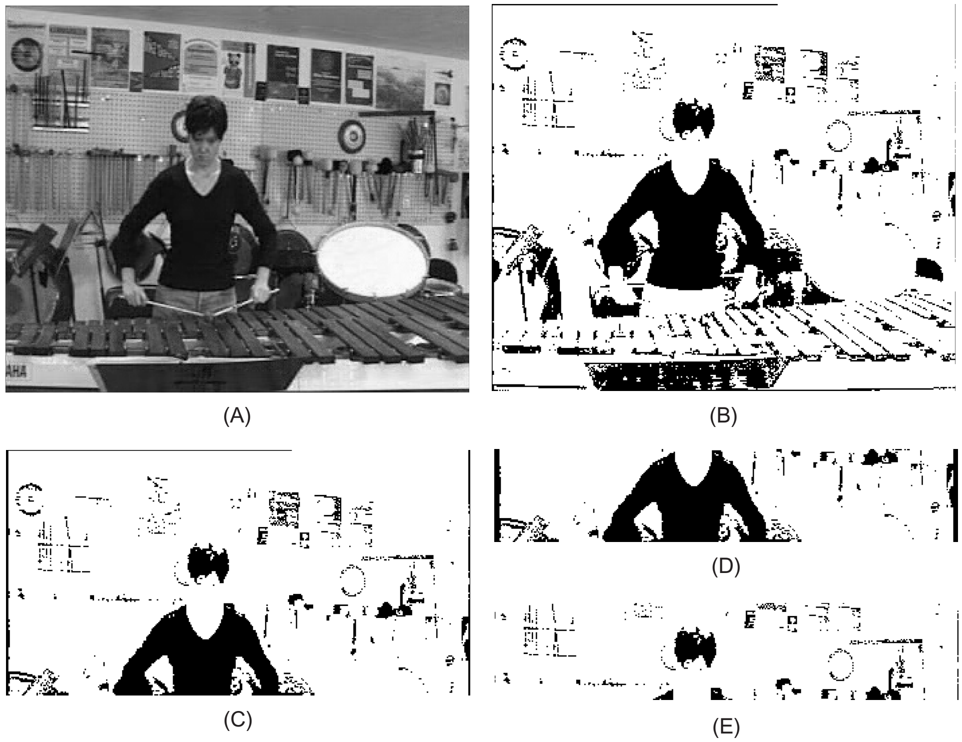

FIGURE 1. Original (top; panel A) and filtered video images exemplifying the four viewing conditions used in the test: full (B), nohands (C), torso (D), and head (E).

The participants were also asked to rate the perceived movement character. Four movement cues were selected, taking into account that they: (a) should describe the general motion patterns of the player (not specific to any part of the body), (b) have a correspondence in musical cues, and (c) reflect characteristics related to the emotional content of the performance rather than the basic transitions required to play the piece. Since the different viewing conditions displayed different parts of the player, specific movement descriptions such as arm direction, head, rotations etc. could not be used.

The cues were (with their musical counterpart in parentheses): Amount (sound level), Speed (tempo), Fluency (articulation), and Regularity (tempo variations). The ratings of the cues were carried out using bipolar scales, presented as horizontal lines with seven markers between the endpoints: "none"–"large" (Amount), "slow"–"fast" (Speed), "jerky"–"smooth" (Fluency), and "irregular"–"regular" (Regularity). For the analysis the ratings were coded as integers from 0 to 6.

The assumptions were that Amount would correspond to an overall measure of the physical magnitude of the movement patterns, Speed to the overall number of movement patterns per time unit, Fluency to the smoothness of movement patterns, and Regularity to the tempo variation in movement patterns over the performance.

The 32 video clips were presented on a computer screen and rated individually. For each participant a command-file automatically opened the clips in Windows Media Player in a randomized order. Each clip could be viewed as many times as the participant liked, but once the window for a specific clip had been closed, the next clip started automatically and the participant could no longer go back to rate the previous one. In order to get the participants acquainted to the procedure, a pre-test was run. During the pre-test the participant was able to see and rate examples of the four viewing conditions and different motional/movement characteristics.

Measure of Achievement

The use of rating adjectives on individual scales results in many values for each stimulus, presentation, and participant. It thus may be useful to calculate a summary statistic that includes all ratings, for the purpose of including the independent factors in one analysis of variance.

Previous examples of how to combine several rated scales into one measure, with the objective of describing emotional communication, can be found in the literature. For example, Juslin (2000) defined *achievement* as the point-biserial correlation (r) between the performer's expressive intention and the listener's rating. This was one of the measures used in the Brunswikian lens model suggested by Juslin for modeling the communication of emotion in music performance. Recently, Resnicow, Salovey, and Repp (2004) calculated emotion recognition scores (E) for each participant by dividing the rating of the relevant emotion by the sum of all four emotion ratings and subtracting a "baseline score" for a normal performance.[1]

One drawback of these estimations is that they do not consider the absolute magnitude of the ratings as will be shown below. Instead, we suggest using the covariance (*Cov*), between intention and rating. The *Cov* reflects both the absolute magnitude of the rating, as well as the ambiguous and/or confused cases. The correlation can be seen as a normalized covariance, where r_{xy} represents the Cov_{xy} divided by the standard deviations for x and y. However, such normalization may result in peculiar behavior when applied to individual ratings of a few adjectives. One particular problem we have found is that r is undefined when all ratings are equal, yielding a standard deviation of 0. An alternative normalization strategy is to normalize relative to the best possible case rather than relative to the actual spread in the data.

Thus, we define the *achievement* (A) as the covariance between the *intended* ($\boldsymbol{x}$) and the *rated* ($\boldsymbol{y}$) emotion for each video presentation, divided by a constant C. Both $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors that consist of four numbers representing Fear (F), Anger (A), Happiness (H), and Sadness (S). The intention vector $\boldsymbol{x}$ was dichotomously coded with the value 1 for the intended emotion and 0 for the other emotions. For the intended emotion Anger $\boldsymbol{x}$ = [F A H S] = [0 1 0 0], the maximum achievement would be for a rating of $\boldsymbol{y}$ = [F A H S] = [0 6 0 0]. The achievement $A(\boldsymbol{x}, \boldsymbol{y})$ for a specific presentation is defined as:

$$A(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{C} Cov(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{C}\frac{1}{N-1}\sum_{i=1}^{N} \overbrace{(x_i-\bar{x})}^{\text{intention}}\overbrace{(y_i-\bar{y})}^{\text{rating}}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are arrays of size N (in our case $N = 4$), and $\bar{x}$ and $\bar{y}$ are the mean values across each array. C is a normalization factor to make the "ideal" achievement equal to 1. Given that $\boldsymbol{x}$ can only take the values 0 and 1, and $\boldsymbol{y}$ can be integer values between 0 and 6, $C = 1.5$ in all cases.

A comparison of the methods for different cases of intentions and ratings is shown in Table 1. The table shows the achievement A, the correlation coefficient r, and the emotion recognition score E between the intention vector for anger $\boldsymbol{x}$ = [F A H S] = [0 1 0 0] and different $\boldsymbol{y}$ values. As seen in the table, A reflects the magnitude of the ratings. A rating of $\boldsymbol{y}$ = [0 6 0 0] gives a higher A value than $\boldsymbol{y}$ = [0 3 0 0], whereas r and E will generate the same value for many different responses (top four rows).

In cases of ambiguity between two emotions (two emotions rated equally high), r will be similar regardless of the "intensity" of confusion. A, on the other hand, gives high values if the two ambiguous emotions are rated high, and low if they are rated low (compare cases for $\boldsymbol{y}$ = [0 6 6 0] and $\boldsymbol{y}$ = [0 3 3 0]). Note that an equal rating of all four emotions, e.g., a rating vector $\boldsymbol{y}$ = [1 1 1 1], (with a standard deviation of 0), does not yield any numerical value for r. Therefore r is lower in discriminant validity and makes a weaker candidate for a measure. The emotion recognition score E always yields numerical values, but the value is the same value for many different cases.

A negative achievement would mean that the intended emotion is confused with other emotions, and zero is obtained when all possible emotions are ranked equally. We assume that an achievement significantly larger than zero (as tested in a two-tailed t-test) implies that the communication of emotional intent was successful. Resnicow et al. (2004) also defined a successful communication when E is significantly larger than zero. However, as E does not take any negative values for confusing cases, A is a more diagnostic measure.

The major part of the data manipulation and statistical analysis was performed using the open source software R (R Development Core Team, 2006). For the analysis of variance SPSS was used.

[1]In the following comparison we will disregard the subtraction of the baseline. Being a constant value for each specific intention, the baseline does not change the sensitivity behavior of the emotion recognition score as compared to the other measures discussed.

TABLE 1. Comparison between achievement (*A*), point-biserial correlation (*r*), and the emotion recognition score (*E*) calculated for combinations of the intention vector for Anger x = [F A H S] = [0 1 0 0] and different rating vectors y.

	y = [F A H S]	A	r	E
Intention correctly identified	0 6 0 0	1.00	1.00	1.00
(ranked highest)	0 1 0 0	.17	1.00	1.00
	0 3 0 0	.50	1.00	1.00
	2 3 2 2	.17	1.00	.33
Ambiguous or equal ranking	0 6 6 0	.67	.58	.50
	0 3 3 0	.33	.58	.50
	1 1 1 1	.00	—	.25
Confusion or non-successful communication	1 0 0 0	−.05	−.33	.00
	3 0 0 0	−.17	−.33	.00
	6 0 0 0	−.33	−.33	.00
	6 5 5 5	−.05	−.33	.24

Results

EMOTION RATINGS

The results from the emotion ratings can be seen in Figure 2. Each panel shows the mean ratings for the four emotions averaged across the 20 participants and the two performances of each intended emotion. The patterns of the bars show the four viewing conditions: full (horizontally striped), nohands (white), torso (grey), and head (diagonally striped), the 95% confidence intervals are indicated by the vertical error bars. Figure 2 indicates that the player was able to convey three of the four intended emotions to the participants in most viewing conditions. Sadness appears to be the most successfully identified emotion, followed by Happiness and Anger. By contrast, Fear was hardly recognized at all but show ratings evenly spread across the four available emotions. The occasional confusion of Anger with Happiness and vice versa indicates that these two expressions might have some features in common.

The tentative conclusions from Figure 2 were supported by the achievement measure, displayed in Table 2, which shows the mean achievement, averaged across 20 participants and two performances, for each intended emotion and viewing condition. The viewing condition receiving the highest achievement for each specific intention is shown in bold. As seen in Table 2, 13 of the 16 combinations of intent and viewing conditions resulted in successful communication (two-tailed *t*-tests). To investigate the effects of the intended emotions and viewing conditions, the achievement measures were subjected to a two-way (4 Conditions × 4 Emotions) repeated measures ANOVA. The effect of intended Emotion was significant and large, $F(3, 36) = 19.05$, $\eta^2 = .61$, $p < .0001$. The effect of Viewing Conditions, was also significant but generated a smaller effect, $F(3, 36) = 6.98$, $\eta^2 = .37$, $p < .001$. A significant but smaller effect was found for the Viewing Condition × Emotion interaction, $F(9, 108) = 4.36$, $\eta^2 = .27$, $p < .0001$. The main effect of Emotion was clearly due to the low achievement obtained for the intention Fear. A Tukey post hoc test, using pairwise comparison, showed that the Fearful intention received significantly lower ($p < .0001$) achievement than all the other three intentions.

For the main effect of viewing Condition, a Tukey post hoc test showed that the torso and head conditions received significantly lower achievement compared to the full condition ($p < .0001$). No other differences between viewing conditions were significant. These results confirmed the a priori assumption that seeing more of the body of the performer improves the achievement.

What viewing conditions were important to detect a specific emotional intention? Table 2 illustrates the interaction between Emotion and viewing Condition. For the Happy intention the torso received higher value than the full condition. For the Sad intention both the head and the nohands condition received higher achievement than the full condition. The achievement values for Happiness, Sadness, and Anger were for all cases significantly greater than zero, ($p < .01$, uncompensated two-tailed *t*-tests), indicating that the communication for these emotions was successful. For Fear, however, only the full condition received achievement values significantly larger than zero ($p < .05$).

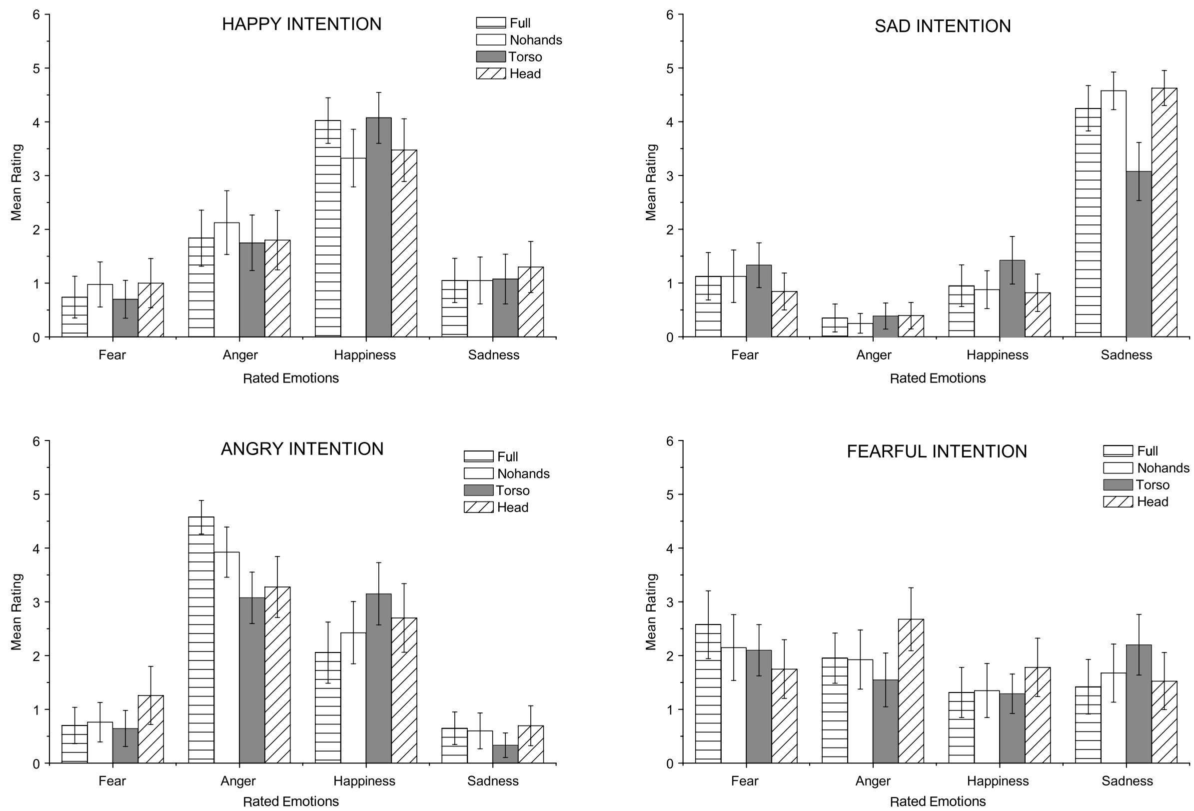

FIGURE 2. Ratings for the four intended emotions and viewing conditions. Each panel shows the mean ratings for the four emotions averaged across 20 participants and the two performances of each intended emotion.

The significant interaction effect was due to differences between conditions for the Sad and Angry intentions. For the Sad intention the head was important for perceiving the intended expression. All conditions where the head was visible (full, nohands, and head) received high achievement values (from .57 to .65 in Table 2), whereas the mean achievement for the torso condition was much lower (.34). A Tukey post hoc test revealed that the only significant effect within the Sad intention was between torso and head ($p < .05$). For Anger, the full condition received the highest achievement (.57), whereas the torso and head conditions were less successful in conveying the intention. The only post hoc effect was between torso and full condition ($p < .05$).

TABLE 2. Mean achievement for the four intended emotions and viewing conditions (full, nohands, torso, and head) averaged across 20 participants and two performances.

Intent	Full	Nohands	Torso	Head	*M*
Happiness	.46***	.32***	**.48***	.35***	.40
Sadness	.57***	.64***	.34***	**.65***	.55
Anger	**.57***	.44***	.27***	.29***	.40
Fear	**.15**	.08	.07	−.04	.07
M	.44	.37	.29	.31	

*$p < .05$ ***$p < .01$

The results for viewing condition in the two-way interaction were somewhat surprising. Initially one would hypothesize that seeing more of the player would provide the participants with more detailed information about the intention. The achievement values would then be ordered from high to low for the three conditions full, nohands and head, and similarly for full, nohands, and torso. Such a “staircase” relation between the viewing conditions was observed in the main effect. However, when achievement scores were broken down by intention this staircase pattern was only found for the Angry condition (see Table 2 and Figure 2).

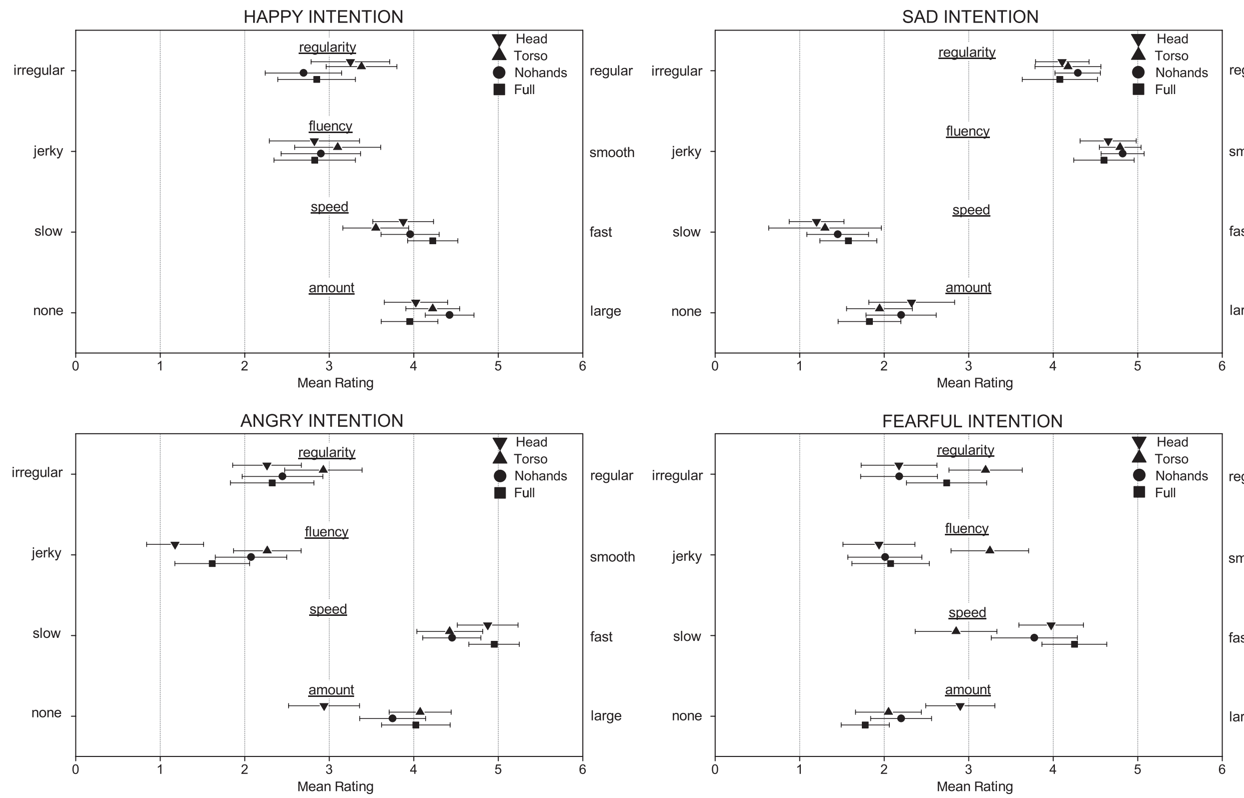

FIGURE 3. Ratings of movement cues for each intended emotion and viewing condition. Each panel shows the mean markings for the four emotions averaged across 20 participants and the two performances of each intended emotion. Error bars indicate 95 % confidence intervals.

MOVEMENT CUES

Figure 3 shows the mean ratings of the movement cues for each intended emotion. The four viewing conditions are indicated by the symbols: full (square), nohands (circle), torso (pyramid), and head (top-down triangle). The movement cues, Amount ("none"– "large"), Speed ("slow"–"fast"), Fluency ("jerky"– "smooth") and Regularity ("irregular"–"regular"), received different ratings depending on whether the intended expression was Happy, Sad, Angry, or Fearful. Note that high ratings corresponded to large amount of movement, high speed, smooth fluency, and regular movements, whereas low ratings corresponded to small amount of movement, slow speed, jerky fluency, and irregular movements.

The intentions Happiness and Anger obtained rather similar rating patterns, explaining part of the confusion between these two emotions. According to the ratings, both Anger and Happiness were characterized by large movements, with the Angry performances somewhat faster and jerkier compared to the Happy performances. The ratings for Fear are less clear-cut, but tend to be somewhat small, fast, and jerky. In contrast, the ratings for the Sad performances display small, slow, smooth, and regular movements.

Table 3 shows the intercorrelations between the movement cues. As expected, they were all somewhat correlated with values ranging from –.62 to .58. The Amount of movement seems to be relatively independent, reflected in the small correlations with the other cues. Speed, Fluency, and Regularity all show relatively medium intercorrelations.

In order to investigate how the rated emotions were related to the rated movement cues, a multiple regression

TABLE 3. Intercorrelations between the movement cues rated in Experiment 1.

Movement cue	Amount	Speed	Fluency	Regularity
Amount	—			
Speed	.26**	—		
Fluency	–.19**	–.62**	—	
Regularity	–.12**	–.44**	.58**	—

**$p < .01$

TABLE 4. Regression analysis beta-weights and semipartial correlations (in italics) for the intentions and rated movement cues in Experiment 1.

Intent	Amount	Speed	Fluency	Regularity
Happiness	.35***	.21***	.06	–.01
$R^2 = .19$	*.34****	*.16****	*.04*	*–.01*
Sadness	–.18***	–.43***	.15***	.08*
$R^2 = .42$	*–.18****	*–.33****	*.11****	*.06**
Anger	.19***	.18***	–.30***	–.16***
$R^2 = .37$	*.18****	*.14****	*–.21****	*–.13****
Fear	–.29***	–.05	–.17**	–.06
$R^2 = .10$	*–.28****	*–.04*	*–.12***	*–.05*

*$p < .05$ **$p < .01$ ***$p < .001$

analysis was performed. Each rated emotion was predicted using the four movement ratings as independent variables. In Table 4 the resulting multiple correlation coefficients (adjusted R^2), the standardized beta-weights, and the semipartial correlations are presented for each rated emotion. The overall multiple correlation coefficients yielded rather low values in terms of explained variance, ranging from 10 to 42%. In general, each rated emotion scale contains more low ratings than high ratings, causing the regression to be somewhat poor. Applying multiple correlation on averaged ratings across participants increases the explained variance to between 67 and 92%. However, due to the few cases available (32) in the averaged ratings, the prediction of the beta weights becomes uncertain in this case.

The semipartial correlation *sr* was used to estimate the relative importance of each movement cue (shown in italics in Table 4). It expresses the unique contribution from each independent variable, excluding the shared variance (Cohen, Cohen, West, & Aiken, 2003). According to the table, the cue that was the most important for predicting Happiness was Amount (large, $sr = .34$), followed by Speed (fast, $sr = .16$). Similarly, the most important cues for Anger were Fluency (jerky, $sr = –.21$), Amount (large, $sr = .18$), and to a lesser degree Speed (fast, $sr = .14$), and Regularity (irregular, sr = –.13).

In general, differences among cue ratings for different viewing conditions were small. For the intentions Happy and Sad and partly for Anger, the cue ratings are closely clustered (see Figure 3). Again, the head seems to play a special role. When a rating stands out from the other viewing conditions it is either for the head or for the torso. Since the latter is the only condition where the head is not visible, it can in fact also be related to the movements of the head.

Experiment 2

To further investigate the robustness of the overall communication through musicians' body movements a second experiment was conducted. Specifically, an objective was to investigate the communication of specific emotions in performances on instruments where the sound producing movements are small and intimately connected to the instrument, such as woodwinds. In addition, we wanted to investigate the generalizability of the results in Experiment 1 by increasing the number of performers and pieces.

Method

STIMULUS MATERIAL

Two professional woodwind players, one soprano saxophonist and one bassoon player, were asked to perform four short musical excerpts with different emotional intentions. Originally, two baritone saxophonists were also recorded. However, an inspection of the recordings showed no sign of expressive gestures in their movements. The movement amplitude was so small that it was considered too diffcult for observers to rate them according to expressive content.

Four melodies were used for the performances: 1) Berwald's String Quartet No. 5, C major, bars 58 to 69; 2) Brahms' Symphony Op. 90 No. 3 in C minor, first theme of the third movement, *Poco allegretto*; 3) Haydn's Quartet in F major for strings, Op. 74 No. 2, theme from first movement; and 4) Mozart's sonata for piano in A major, K331, first eight bars. Unlike the piece used for the performances in Experiment 1, which was selected to be of a neutral character, these melody excerpts were chosen so as to vary the compositional/structural contribution to the emotional expression (cf., Gabrielsson & Lindström, 2001).

Before the recordings, the players received the scores together with written instructions to prepare performances with different emotional expressions. All four melody excerpts were to be performed portraying 12 emotional expressions (not all used in this particular experiment). The players were instructed to perform the different excerpts so as to communicate the emotions to a listener as clearly as possible. The instructions made it clear that the emphasis was on the musical interpretation of the emotion. After recording the 12 intentions, an "indifferent" performance was recorded.[2]

[2] The "indifferent" performances were primarily recorded for the purpose of audio analysis. For audio, the melody itself may contain some "baseline expression" (see e.g., Resnicow et al., 2004) whereas there is no such baseline in terms of movement. Instructed to perform indifferent, the players simply did not move more than the bare necessity.

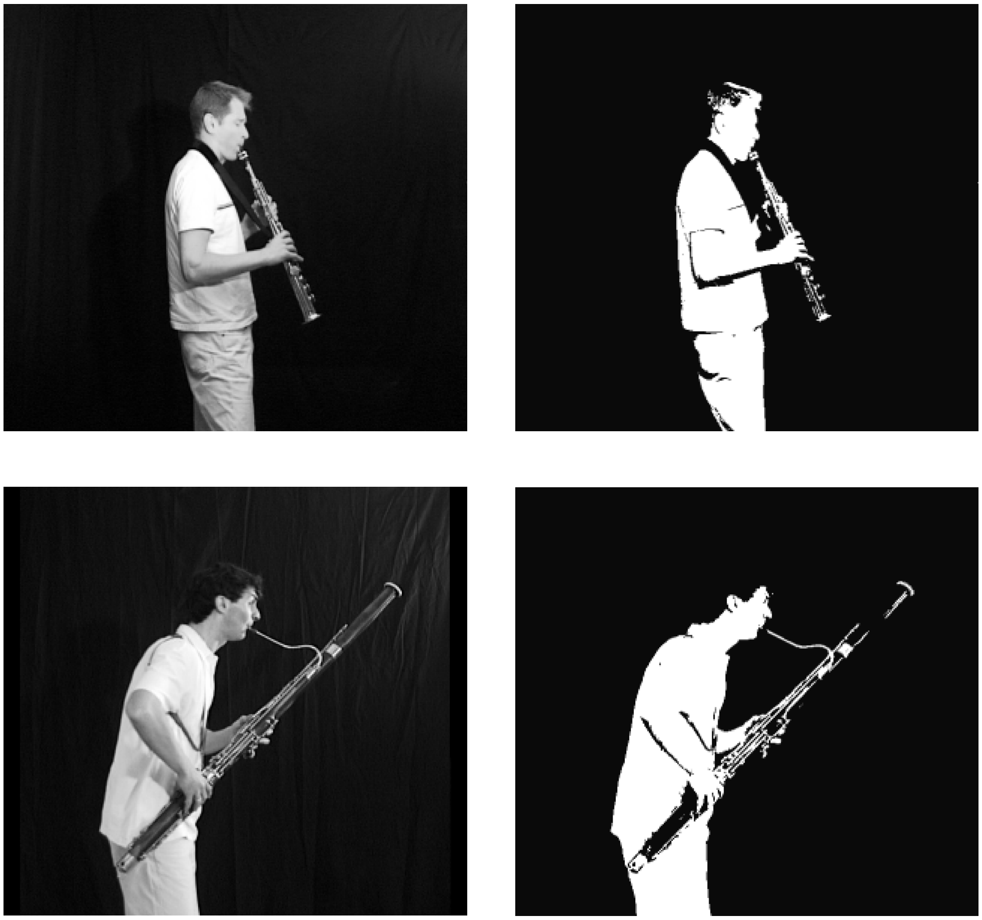

FIGURE 4. Original and filtered video images exemplifying clips of the woodwind players: saxophone (top two pictures) and bassoon (bottom two).

The movements were recorded using the same digital video camera as in Experiment I.[3] The camera was placed on a stand at a fixed distance on the players' right side. To enhance the contrast between the player (who was asked to dress in light colors) and the background (black curtains), additional spotlights and short shutter time for the camera were used. From the 12 emotional expressions recorded, the performances of Happiness, Sadness, Anger, and Fear were selected as video stimuli.

[3]As the purpose of the recordings was to provide stimuli for several investigations with different purposes, the recording procedure differed from that in Experiment 1. Specifically, both video and high-quality audio recordings of the performances were made. The players were informed that both audio and movements could be participant to analysis but not in which way. The audio recordings were not used in the experiment reported here.

The editing of the video clips was similar to that in Experiment 1. This time, however, no differing viewing conditions were generated. The reason was that wind instrumentalists are intimately connected to their instrument with relatively small sound producing movements (as compared to percussionists). Examples of original and filtered video frames showing each of the two players can be seen in Figure 4. In total 32 (4 Emotions × 2 Players × 4 Excerpts) video clips were generated. The duration of the video clips varied between 9 and 46 s.

PARTICIPANTS

A total of 20 participants (10 male and 10 female) volunteered to participate in the experiment. The participants were between 23 and 59 years old ($M = 31, SD = 8$)

with varying amounts of music training. Five participants reported that they did not play any instrument. Ten of the participants had several years of experience with one or more instrument and played regularly. Participants recruited from outside the department received a small compensation for their participation. None of the participants had participated in Experiment 1.

PROCEDURE

The participants were asked to rate the same parameters as those in Experiment 1. The emotional content in the video clips was rated on a scale from 0 (“nothing”) to 6 (“very much”), for the four emotions Fear, Anger, Happiness, and Sadness. The ratings of movement character were carried out on bipolar scales for each of the cues Amount (“none”–“large”), Speed (“slow”–“fast”), Fluency (“jerky”–“smooth”), and Regularity (“irregular”– “regular”). A difference from Experiment 1 was that the rating scales were not restricted to integers, but could take any value between 0 and 6.

The 32 video clips were presented on a computer, using a custom-made graphical user interface, and were rated individually. The stimuli clips were presented in two blocks, one block with all saxophone clips, randomized for each participant and session, and another block with all bassoon clips. Half of the participants started rating the block with saxophone clips first and half started by first rating the bassoon clips. Each clip was repeatedly played until the participant had rated all parameters. It was not possible to go on to rate a new clip until all parameters had been rated, and once the next clip was started the participant could no longer go back to rate the previous one. Similarly to Experiment 1 a pre-test was run, allowing participants to see and rate examples of the two players and different emotional/ movement characteristics.

Results

EMOTION RATINGS

The results from the emotion ratings for the two performers can be seen in Figure 5. Each panel shows the mean ratings for the bassoon player (striped bars) and saxophonist (grey bars) and four emotions, averaged across the 20 participants and the four musical excerpts. The vertical error bars indicate the 95% confidence intervals. Comparing Figure 5 to Figure 2, the results are very similar: the intentions Happiness, Sadness, and Anger were communicated to the participants, whereas Fear was not. In general, however, the ratings were lower compared to those in Experiment 1. There seemed to be less confusion between Anger and Happiness for the woodwind performers than for the marimba player, suggesting some differences in movement cues.

As in Experiment 1, achievement was calculated for each of the presented video clips (see equation). Table 5 shows the mean achievement values for the two performers, four excerpts, and each intended emotion. The performance receiving the highest achievement for each specific intention is shown in bold. The mean achievement for the Fearful intention (–.05) is considerable lower than the intentions Happiness (.34), Sadness (.43), and Anger (.34). The significance levels in the table indicate the outcome of two-tailed *t*-tests (not compensated for multiple comparisons).

The achievement measure was subjected to an analysis of variance to investigate the effects of the intended emotions and musical excerpts. The three-way (4 Excerpts × 4 Emotions × 2 Performers) repeated measures ANOVA showed a large main effect for Intended Emotion, $F(3, 57) = 42.06$, $\eta^2 = .69$, $p < .0001$, and significant, but smaller, effects of musical Excerpt, $F(3, 57) = 11.54$, $\eta^2 = .38$, $p < .0001$, and Performer, $F(1, 19) = 5.46$, $\eta^2 = .22$, $p < .05$. All two-way interactions also showed significant effects: Excerpt × Emotion, $F(9, 171) = 6.65$, $\eta^2 = .26$, $p < .0001$, Emotion × Performer, $F(3, 57) = 12.61$, $\eta^2 = .40$, $p < .0001$, and Excerpt × Performer, $F(3, 57) = 7.21$, $\eta^2 = .28$, $p < .0001$.

Similarly to Experiment 1, the main effect of emotion was clearly due to the low achievement obtained for the intention Fear. A Tukey post hoc test showed that the achievement values for the Fearful intention were significantly lower ($p < .0005$) compared to those of the other three intentions.

The main effect of performer was due to slightly higher achievement values for the bassoon player compared to the saxophonist (see Table 5). The significant interaction between Performer and Emotion was mainly due to the low achievement for the intention Happiness for the saxophonist, who was rated sad to a higher degree (cf., Figure 5). The Happy intentions performed by the saxophonist received significantly lower achievement values than all other cases ($p > .001$), except his own Anger performances.

The main effects due to musical excerpt were expected, considering that the excerpts were chosen on the basis of their different emotional characteristics. A Tukey post hoc test revealed that the Brahms excerpt received significantly lower achievement compared to the other excerpts ($p < .05$). In addition, the Haydn excerpt

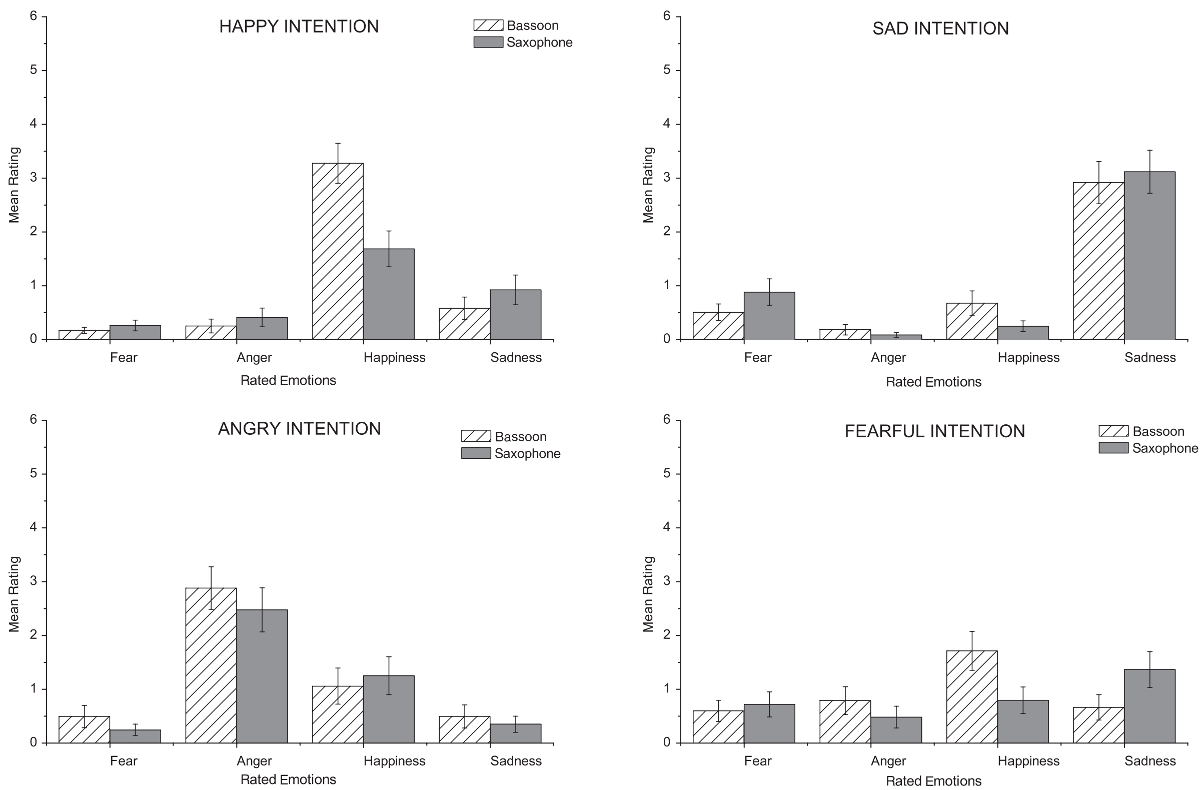

FIGURE 5. Ratings for the four intended emotions and two instrumentalists; bassoon player (striped bars) and saxophonist (grey bars) averaged across 20 participants and the four musical excerpts.

received significantly higher achievement values than the Mozart excerpt ($p < .05$).

Were any of the excerpts better at communicating a certain emotion? The largest difference between intentions for one particular excerpt was found for Mozart. For this excerpt, Happiness received significantly lower achievement than Sadness and Anger ($p < .001$). For the Brahms excerpt Anger received significantly lower achievement than the Sad intention ($p < .01$). This corresponds well to the inherent character of these excerpts, the Brahms being a slow, minor tonality melody (Gabrielsson & Lindström, 2001). By contrast, the Berwald and Haydn excerpts displayed no significant differences between Happy, Sad, and Angry intentions.

TABLE 5. Mean achievement for the bassoon and saxophone performers, excerpts and intended emotion.

Intent	Berwald	Brahms	Haydn	Mozart	*M*
		Bassoon			
Happiness	**.68*****	.44***	.63***	.22**	.49
Sadness	.33***	.39***	.48***	.44***	.41
Anger	.44***	.16	**.48*****	–.39***	.37
Fear	–.04	–.03	–.07	–.17***	–.24
M	.35	.24	.38	.22	
		Saxophone			
Happiness	.17**	.15*	.33***	.12	.19
Sadness	.47***	.30***	.44***	**.60*****	.45
Anger	.24***	.12	**.48*****	.41***	.31
Fear	.00*	–.13***	**.02**	.01	–.03
M	.22	.11	.31	.29	

*$p < .05$ **$p < .02$ ***$p < .001$

MOVEMENT CUES

The mean ratings for the two performers can be seen in Figure 6. As seen in the figure, the movements for the bassoon player (squares) and the saxophonist (circles) were rated similarly in many cases. The movement ratings of the intended emotions also resemble those in Experiment 1, especially for Anger and Sadness (cf., Figure 3).

Table 6 shows the intercorrelations between the rated movement cues for the two performers in Experiment 2. For the saxophonist, the movement cues were all somewhat correlated, with values ranging from –.80 to .58.

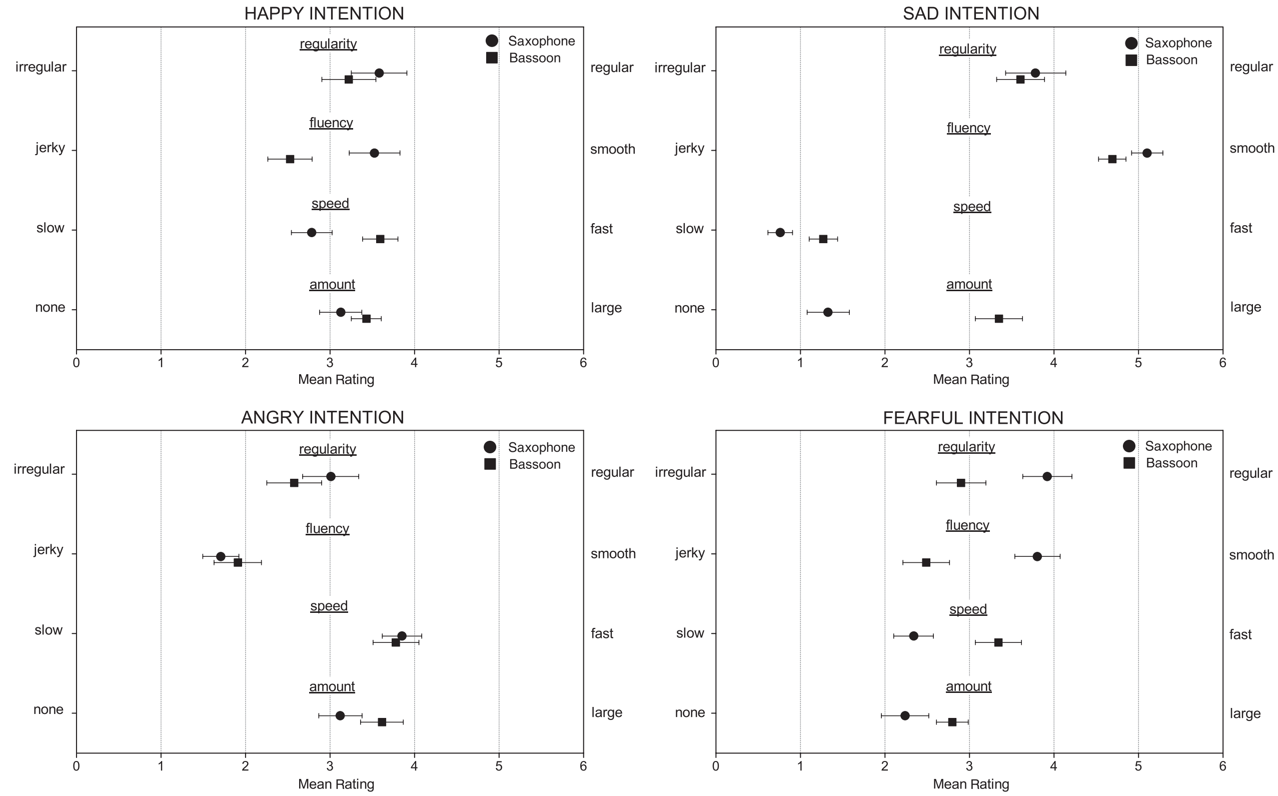

FIGURE 6. Mean ratings for the movement cues and two performers in Experiment 2, averaged across four excerpts and 20 participants. Error bars indicate 95 % confidence intervals.

For the bassoon player three of the cues, Speed, Fluency, and Regularity, are correlated with values ranging from –.77 to .20. Amount, however, is independent from the other cues (similar to Experiment 1, see Table 3), suggesting differing movement characteristics for each player.

TABLE 6. Intercorrelations between the movement cues for the two performers rated in Experiment 2.

Intent	Amount	Speed	Fluency	Regularity
		Bassoon		
Amount	—			
Speed	–.01	—		
Fluency	.02	–.77**	—	
Regularity	–.07	–.14**	.20**	—
		Saxophone		
Amount	—			
Speed	.58**	—		
Fluency	–.43**	–.80**	—	
Regularity	–.18**	–.28**	.41**	—

**$p < .01$

As in Experiment I, the relation between *rated emotion* and *movement cues* was investigated through a multiple regression analysis. Table 7 displays the resulting multiple correlation coefficients, the standardized beta-weights, and the semipartial correlations (*sr*, in italics) for the two performers in Experiment 2. In general, the results for the bassoon player and the saxophonist are similar to those for the marimba player in Experiment 1. The explained variance was also similar to Experiment 1, with values ranging between 8 and 47%.

Compared to the results for the marimba player, there were fewer cues identified per emotion for the wind players. None of the rated emotions display more than two significant movement cues. Also, the overlapping of movement characteristics between Anger and Happiness seems to be absent for these two performers.

According to the table, the most important cue for predicting the rated Anger is Fluency (jerky, sr = –.18 and –.25). Neither Amount, nor Speed, was significantly contributing to Anger ratings. For Happiness a difference

TABLE 7. Regression analysis beta-weights and semipartial correlations (in italics) for the intentions and rated movement cues in Experiment 2.

	Movement cues			
Intent	Amount	Speed	Fluency	Regularity
		Bassoon		
Happiness	.04	.36***	–.03	.01
$R^2 = .13$	*.04*	*.23****	*–.02*	*.01*
Sadness	–.02	–.49***	.24***	.00
$R^2 = .47$	*–.02*	*–.31****	*.15****	*.00*
Anger	.06	.14	–.29***	–.08
$R^2 = .18$	*.06*	*.09*	*–.18****	*–.08*
Fear	–.12*	–.12	.04	–.02
$R^2 = .02$	*–.12**	*–.08*	*.02*	*–.02*
		Saxophone		
Happiness	.24***	.19	.05	–.01
$R^2 = .11$	*.20****	*.10*	*.03*	*–.01*
Sadness	–.08	–.37***	.29***	–.09
$R^2 = .41$	*–.07*	*–.20****	*.16****	*–.08*
Anger	.01	.17	–.44***	.00
$R^2 = .33$	*.01*	*.09*	*–.25****	*.00*
Fear	–.23***	.10	.24*	–.08
$R^2 = .08$	*–.18****	*.05*	*.13**	*–.07*

*$p < .05$ ***$p < .001$

between the two performers' movement cues can be seen. For the bassoon player, the most important cue to predict Happiness was Speed (fast, *sr* = .23), whereas Amount (large, *sr* = .20) was important for the saxophonist. The cues most important to predict Sadness were Speed (slow, *sr* = –.31 and *sr* = –.20) together with Fluency (even, *sr* = .15 and .16). Fear was characterized by Amount (small, *sr* = –.12 and –.18) and, in the case of the saxophonist, also Fluency (even, *sr* = .13). Considering the very low ratings for Fear, however, its characterization is questionable.

Results Across Performers

The combined result of Experiment 1 and 2 is shown in Figure 7. The figure shows mean ratings for the four rated emotions averaged across all players, participants, and cases (only including the full viewing condition in Experiment 1). As seen in the figure, the intentions Happiness, Sadness, and Anger received many ratings according to the intention, whereas Fear does not appear to be recognized at all.

The results were also examined by transforming the participants' ratings into percent correct scores,

FIGURE 7. Mean ratings for the four rated emotions averaged across all players, participants and cases in Experiments 1 and 2. Each bar is based on 200 values, only the full viewing condition in Experiment 1 has been included.

as is commonly done in other studies. The transformation was done in a strict fashion, meaning that only the ratings where the intended emotion received the highest rating were considered as "correct." The total percentage of correct identifications of the four intended emotions, across both experiments, were 68% for Happiness, 81% for Sadness, 74% for Anger, and 23% for Fear (see Table 8). The performer who received the most correct identified intentions varied with emotion. The bassoon player received the highest values for Happiness (81%), and for Sadness the saxophonist received 83% correct, but recognition was high for all three performers for this intention. Only the full viewing condition was considered for the marimba player, who received the highest value for Anger (83%) and Fear (35%). Across all performers,

TABLE 8. Correct identification of the intended emotions in percent for each of the three performers rated in Experiments 1 and 2. The performer receiving most correct identifications for a specific intention is shown in bold.

Intent	Marimba	Bassoon	Saxophone	*M*
Happiness	68	**81**	56	68.3
Sadness	80	80	**83**	81.0
Anger	**85**	72	64	73.7
Fear	**35**	14	19	22.7
M	67.0	**61.9**	55.3	

TABLE 9. Comparison between the movement cues for Experiments 1 and 2, and reported auditory cues expressing emotional intentions.

	Movment cues				Auditory cues			
Intent	Amount	Speed	Fluency	Regularity	Sound lev.	Tempo	Articulation	Tempo var.
Happiness	large * •	fast * ∘			high	fast	staccato	small
Sadness	small *	slow * ∘ •	smooth * ∘ •	regular *	low	slow	legato	final ritard
Anger	large *	fast *	jerky * ∘ •	irregular *	high	fast	staccato	small
Fear	small * ∘ •		jerky * •		low	fast	staccato	large

*marimba ∘ sax • bassoon

however, Fear barely received 23% correct, which was below chance level (25%).

The differences between the different performers in their use of movement cues make a combined multiple regression across performers less suitable. However, by compiling the important movement cues for each player, some general conclusions can be drawn. Table 9 presents the significant movement cues for the three performers in Experiment 1 and 2. As seen in the table, there are general trends for cue utilization across performers. Sadness was characterized by slow and smooth movements, whereas Anger was characterized by jerky movements. Large and somewhat fast movements characterized Happiness. Most important to characterize Fear were small movements, but jerky Fluency also had some importance.

For comparison, four auditory cues that are important for identifying emotions when listening to music performances are shown to the right in Table 9: sound level, tempo, articulation, and tempo variability (selected from Juslin, 2001). Note the close correspondence between the movement and audio cues for the four emotions. For each rated emotion, the important cues and their corresponding audio counterparts all change in the same direction with the exception of Regularity.

Conclusions and Discussion

In this study we have shown that it is possible for a musician to convey specific emotions using body movements only. Additional objectives for the study were to investigate: 1) the overall communication of each specific emotion, 2) how the communication is influenced by which parts of the player are shown, 3) whether movement cues can be used to describe this communication, and 4) to what extent different performers affect the communication. Our results indicate that:

1. The intended emotions Sadness, Happiness, and Anger were successfully conveyed, whereas Fear was not.
2. The identification of the intended emotion was only slightly influenced by the viewing condition, although in some cases the head was important.
3. Rated movement cues could be used to characterize the different emotional intentions. Anger was primarily associated with jerky movements, Happy with large and somewhat fast movements, Sadness with slow and smooth movements, and Fear with small and somewhat jerky movements. However, since the communication of Fear failed, its characterization is questionable.
4. Although there are differences in ratings between performers they also seem to share several movement cues that could be of importance to observers.

Overall Communication of Intended Emotion

The achievement values of the communicated emotional intentions in our experiment correspond well to earlier studies of expressive movements in dance performances. Given the choices of different emotions, participants identified the intentions well above chance in many cases (e.g., Boone & Cunningham, 1998, 2001; Dittrich et al., 1996; Walk & Homan, 1984). The emotion most correctly identified differed between studies, but Anger, Sadness/Grief and Happiness/Joy generally received a large portion of correct responses.

Overall, the mean achievement for Sadness was higher compared to the other three intentions in the present study. Sadness having been the most successfully conveyed intention is not surprising, considering that children as young as four years old are able to identify Sad performances (Boone & Cunningham, 1998), and also produce the relevant cues that allow adult observers to identify Sadness above chance level (Boone & Cunningham, 2001). The ability to identify and portray Fear, Anger, and Happiness appears later, from the age of five.

Other researchers also have found confusion between Happiness and Anger. Dittrich et al. (1996) reported that when presenting observers with point-light dance performances, Anger was likely to be mistaken for Joy, and vice versa. When the dances were showed in normal lighting, however, the confusion appeared between Joy and Surprise. Also Sörgjerd (2000) found that Anger and Happiness were often confused with one other, and suggested a possible explanation to be the presence of high activation in both these intentions.

Fear was the only intention in the present study for which the communication failed. This is not in correspondence with other studies. Sörgjerd (2000) found no significant differences in the communication of Fear compared to Happiness, Sadness, and Anger. Although the percent correct identifications were worse for Fear compared to the other intentions in the watch-only condition, the intentions Sadness and Anger received only marginally better scores. However, it is not clear whether facial expressions were visible for the observers in her study.

There are several possible explanations for the communication failure of Fear in the present study. First, it is possible that the players were unable to convey Fear in all modalities, not only the visual. It may well be that the performers' interpretation of fear was closer to anxiety or nervousness. Music performances are capable of expressing a wide range of emotional expressions. Sounded performances can communicate both expressive intentions inherent in the original score as well as those added during the performance (Juslin & Sloboda, 2001). Although it seems unlikely that three performers would all be unable to express an emotion that has been successfully identified in a number of listening tests, this possibility cannot be excluded. A replication of Experiment 2 with auditory stimuli only would answer this question.

Second, even if the musicians were playing the sounded notes according to the intention Fear, it is possible that they repressed movements more for this intention than for others. If the way the players chose to interpret the intent Fear involved "freezing" behavior there would not be much movement for observers to rate. Also, performers that experience stage fright must learn to tackle possible physiological signs of their distress in order to deliver a musically acceptable performance. In fact, we are not likely to consider performances that show signs of anxiety or fear to be good or professional.[4] It therefore seems plausible that movement cues revealing any anxiety or fear are consciously suppressed by the performer.

Along similar lines, a third possibility is that the players were able to express Fear both auditory and visually, but the observers did not accept the emotion as being plausible. Although some music is composed with the explicit intention of inducing Fear (e.g., Herrmann's original film music to "Psycho") very few listeners expect a professional musician to express real Fear while performing. It is simply not part of what we expect music performances to be about unless it takes place within a more narrative context (e.g., opera).

It could well be that observers have difficulty imagining someone who is frightened to stay on stage and perform. It is plausible that the failing communication of Fear is related to what emotions the observers *expect* to see, rather than how well the intended emotions were portrayed. Some support for this was seen during a pilot test with the recorded stimuli for Experiment 2. In the pilot test participants rated all 12 recorded intentions. From the result of the pilot test it was evident that participants were more likely to consider positive emotions such as Love, Pride, and Contentment as expressed, whereas the negative emotions Disgust, Jealousy, and Shame rarely received any ratings at all. It could be that by eliminating the information from facial expression observers are less disposed towards rating emotional intentions not considered likely to occur.

Movement Cues

A good correspondence was found between how the selected movement cues and their audio counterparts varied for different intentions (see Table 9). This supports the correspondence between movement and audio cues and, once again, the intimate coupling motion—music.

All three performers used slow movements to communicate Sadness, which affected the duration of the video clips for this particular intention. Similar to the

[4]A somewhat ironic fact considering the huge proportion of professional musicians reporting to suffer from performance anxiety (see e.g., Fishbein, Middlestadt, Ottati, Straus, & Ellis, 1988).

essential role of tempo in music performance, the connection between velocity (Speed) and duration of movement can be important for identifying the intention. Paterson et al. (2001) found that manipulating the durations of Angry, Neutral, or Sad lifting and knocking movements had an effect on observers' ratings. There were clear changes in the classification and intensity ratings for all three intentions. Manipulated Angry movements were, however, seldom categorized as Sad or Neutral. Paterson et al. concluded that movement velocity has a role in perceiving intent/affect, but that there are other important properties that are not controlled by velocity.

The cue Regularity did not appear to be as strong as the other cues. Also, comparing cues for audio and movements in Table 9, Regularity is the only movement cue that does not seem to correspond to its auditory counterpart (tempo variability). This could be an indication that the observers did not rate the cue Regularity as was intended.

The multiple correlation coefficients from the regression analysis in Tables 4 and 7 exhibit rather low explained variance (R^2). When applied on average ratings over participants, R^2 values increase and become similar to those obtained in studies using regression analysis to predict emotion ratings from audio cues (e.g., Juslin, 2000). In the present experiment, the participants quantified the cues, whereas in the music performance experiments the cues were measured from the performances. However, the cue estimations are quite consistently appearing in a narrow range for each intended emotion, as illustrated in Figures 3 and 6. A possible explanation for low explained variance can be due to a non-linearity of the estimated cues relatively the emotion ratings. For example, rated Fluency for the Happy intention (see Figure 6) is in the middle of the range whereas Fluency ratings for the angry and sad intention are in the low and high range, respectively. Assuming that the emotion ratings correlate with the intentions, it follows that a linear model will fail to provide significance for Fluency in this case. This is supported by the relatively lower R^2 values for Happiness compared to Sadness and Anger both in Experiment 1 and 2. In their extensive overview of studies of emotional expression in music and speech, Juslin and Laukka (2003) acknowledge a need for three different cue regions in some cases. Thus, to find a non-linear mapping between cues and emotions would be an interesting topic for future studies. One possible candidate that uses different cue regions is the fuzzy mapping method developed for real time emotion recognition (Friberg, 2005).

Effects of Viewing Condition (Experiment 1)

The identification of the intended emotion was only slightly influenced by the viewing condition, although for the Sad intention it was evident that the movements of head provided important cues for identifying the intention correctly. The viewing condition where the head was not visible (torso) obtained lower ratings than the other conditions.

A possible explanation could be that for the Sad intention there is a specific cue from the player visible in the head only. Informal visual inspections of the head stimuli suggested that the head was mostly turned downwards for the Sad performances. Also, there were less and slower movements in the vertical direction compared to the other intentions.

One explanation for the relatively small effect of viewing condition could be that the viewer is able to imagine the non-visible parts of the body. The clips showing only part of the player could have been judged from the imagined movements of the invisible parts. In point-light studies, where sometimes extremely limited information is available for the observer, the ability of "reconstructing" the missing parts could be a strategy when judging what is seen (see e.g., Davidson, 1994).

Another possible explanation could be that the edited clips for the different viewing conditions sometimes interfered with each other. For example, the clips that were edited to display the head often included part of the shoulders. In the condition torso the shoulders were present at all times, but sometimes parts of the player's inclined head would also be visible. The proportion of video frames in the torso condition in which the head was partially visible was about 3% for the Sad performances, 0% for Anger, and about 10% for Happiness. For the intention Fear the second performance revealed the head to a great extent towards a final tremolo in the performance, resulting in a higher proportion of frames showing possible head cues (11 and 27% for the two performances respectively). Similarly, for large stroke movements the mallets could occasionally be visible in the nohands and torso conditions. This usually occurred for only one to two frames at a time.

In our study we did not include a condition with only the hands of the player visible. Davidson (1994) found that the hands provided no information about the expressiveness in the piano performances. A comment from one of the participants illustrates this: "I suspect I'm rating this wrist performance as highly expressive just because there is plenty of action" (Davidson, 1994). From measurements on the movements of the same pianist, Davidson also reported that there were small differences in the extent of hand movements between the different performance conditions. That is, the difference between the ranges in the vertical or horizontal direction was about the same regardless of the performance condition. By comparison, the head movements in both vertical and horizontal directions showed significant differences between performance conditions. Differences were especially evident between deadpan and projected performances.

Effects of Performer

There were individual differences regarding which movement cues the three players used to express the emotional intentions. For instance, the most important cue to characterize Happiness was fast movements for the bassoon player, large movements for the saxophonist, and both large and fast movements for the marimba player.

The somewhat different movement characteristics in the players' performances were an expected result. Not all performers use expressive movements distinctively or in a way that is interpretable by an observer. In fact, many musicians may not appear to move at all. Two of the saxophonists originally recorded displayed so little movement that it was difficult to rate their cues. Including these videos with the others in Experiment 2 may have provided interesting data, but a total of 64 video clips would have rendered test sessions that were too long.

Why is it that some musicians move more than others? Wanderley, Vines, Middleton, McKay, and Hatch (2005) reported that most of the clarinetists in their study claimed that excessive movements could interfere with playing. In some cases musicians also consciously train to refrain from "unnecessary" movements. After our recordings, one of the baritone saxophonists, whose movements were not used, confirmed that he had trained to play with no more movements than necessary. For this player, the idea that only the sounded notes are to be communicated to the audience was a matter of principle and a conscious decision. Certainly the movements of a musician have an effect on the audience's percept. Not only do performer movements affect the perception of expressive intention, but also how individual sounds are perceived (see e.g., Saldaña & Rosenbaum, 1993; Schutz & Lipscomb, in press). However, it is possible that some movements can aid the performance also if they are not directly related to the production of notes. For instance, it has been shown that body sway can improve reading performance (Stoffregen, Pagulayan, Bardy, & Hettinger, 2000).

Movements can also become an internalized part of performance without being obvious to an observer, or even executed. As Davidson and Correia (2002) point out, many of the musicians in Davidson's studies stated that their musical ideas were closely related to repetitive whole-body motion. Davidson and Correia suggest that this imagined sense of motion could mean that musical expression can be internalized as well as externalized. The German pedagogue Truslit had similar ideas and likened musical motion to an "invisible, imaginary dance" (Repp, 1993).

Some support for the internalization of musical expression can be found in the literature. Wanderley (2002) reported that the clarinetists participating in his study reproduced some movement patterns also when were told not to move at all, although with much reduced amplitude (see also Wanderley et al., 2005). A study of expressive singing also supports that performers' expressive behavior are not easily suppressed (Sundberg, Iwarsson, & Hagegård, 1995). In the study, a singer asked to perform in a non-expressive way still made the same types of deviations in timing and sound level, only to less extent.

Influence of the Instrument

In this study we have investigated three soloists performing on different instruments. Admittedly this is a small sample, but nevertheless it could be of some interest to discuss the possible meaning of the body language of percussion players in relation to wind players. A wind player is "attached" to the instrument during playing. To raise the head means that the instrument must follow, and vice versa. A percussionist, however, does not *need* to change movement pattern of the head when performing with different expressive intentions. Some horizontal transitions of the body are necessary when playing the marimba, since the player moves along the instrument. The player also has to read the score and check the positions of the mallets, and this will also enforce some movement of the head. However, there seems to be no reason why the movement cues would differ to such a degree between the intended expressions. Is it possible that the overall body language somehow could be helpful in expressive playing? If so, to what extent could the type of instrument explain differences in the player's movements when performing?

For certain instruments, the sound production movements and visual communicative movements are closely linked. String players and percussionists are good examples of players whose movements closely reflect what they are playing (Askenfelt, 1989; Dahl, 2000, 2004). Percussion playing in general uses large movements, but the player has little control over the tone once it has been initiated. The tone can be shortened by dampening, but not lengthened. While, for instance, players of wind instruments have close control of air stream during the full duration of a tone, the contact time between mallet and drum head is in the order of milliseconds. This implies that whatever dampening or force the percussionist wants to induce has to be part of the gesture from the very beginning. The mallet will strike the drum head (or whatever structure is set into vibration) with the velocity and mass applied through the player's movement, and the same movement gesture will also determine the contact duration.

Pianists have some control over the note length, but otherwise their situation is similar to percussion playing. When the hammer strikes the string in the piano there is no longer any connection between the player's finger on the key and thus the hammer velocity is determined by the history of the key depression. Is it possible that for players of these instruments, gestures in terms of larger movements may not only be important for visualizing intentions but also could play an important role in learning to control the sound production? Further research could reveal whether the movement cues reported here would apply also for other performers and instruments.

Although there are differences in movement patterns between the performers in our study, the similarities seem to dominate. Despite the fact that the required sound generating movements differ considerably between woodwind and percussion players, our results suggest that performers on these instruments share a common body language.

To conclude, our results show that specific emotional intentions in music performance can be communicated to observers through movements only. The intentions Happiness, Sadness, and Anger were successfully conveyed whereas Fear was not. For the marimba performances, the head seemed to play a special role in the communication. Rated movement cues, resembling auditory cues of importance for conveying emotions, could be used to describe the communicated emotions.

Author Note

The authors would like to thank Alison Eddington for the marimba performances, Andrea Bressan and Anders Paulsson for the woodwind performances, all persons participating as subjects in the viewing test, Erwin Schoonderwaldt for help with the Matlab GUI used in Experiment 2, and Anders Askenfelt, Peta Sjölander, and three anonymous reviewers for valuable comments on the manuscript. In addition, the authors would like to thank Jeff Pan for providing helpful consulting on the statistics used.

This work was supported by the European Union (MEGA—Multisensory Expressive Gesture Applications, IST-1999-20410). The collaborations with Andrea Bressan and Anders Paulsson were made possible through the SOCRATES programme: Higher Education (ERASMUS) and the Sweden Tercentenary Foundation sponsored project FeelMe.

Correspondence concerning this article should be addressed to Sofia Dahl and Anders Friberg, KTH School of Computer Science and Communication, Dept. of Speech, Music and Hearing, Royal Institute of Technology, SE-100 44 Stockholm, Sweden; E-MAIL: *sofiadah@kth.se* (Dahl) and *afriberg@csc.kth.se* (Friberg).

References

ABERNETHY, B., & RUSSELL, D. G. (1987). Expert-novice differences in an applied selective attention task. *Journal of Sport Psychology*, *9*, 326-345.

ASKENFELT, A. (1989). Measurement of the bowing parameters in violin playing II: Bow bridge distance, dynamic range, and limits of bow force. *Journal of the Acoustic Society of America*, *86*, 503-516.

BOONE, R. T., & CUNNINGHAM, J. G. (1998). Children's decoding of emotion in expressive body movement: The development of cue attunement. *Developmental Psychology*, *34*, 1007-1016.

BOONE, R. T., & CUNNINGHAM, J. G. (1999). *The attribution of emotion to expressive body movements: A structural cue analysis*. Manuscript submitted for publication.

BOONE, R. T., & CUNNINGHAM, J. G. (2001). Children's expression of emotional meaning in music through expressive body movement. *Journal of Nonverbal Behavior*, *25*, 21-41.

CAMURRI, A., LAGERLÖF, I., & VOLPE, G. (2003). Recognizing emotion from dance movements: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies*, *59*, 213-225.

CLARKE, E. F. (2001). Meaning and specification of motion in music. *Musicae Scientiae*, *5*, 213-234.

CLARKE, E. F., & DAVIDSON, J. W. (1998). The body in performance. In W. Thomas (Ed.), *Composition—Performance—Reception* (pp. 74-92). Aldershot: Ashgate.

COHEN, J., COHEN, P., WEST, S. G., & AIKEN, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New Jersey: Lawrence Erlbaum Associates.

DAHL, S. (2000). The playing of an accent - preliminary observations from temporal and kinematic analysis of percussionists. *Journal of New Music Research*, *29*, 225-233.

DAHL, S. (2004). Playing the accent - comparing striking velocity and timing in an ostinato rhythm performed by four drummers. *Acta Acustica united with Acustica*, *90*, 762-776.

DAHL, S., & FRIBERG, A. (2004). Expressiveness of musician's body movements in performances on marimba. In A. Camurri & G. Volpe (Eds.), *Gesture-based communication in human-computer interaction* (pp. 479-486). Berlin, Heidelberg: Springer Verlag.

DAVIDSON, J. W. (1993). Visual perception and performance manner in the movements of solo musicians. *Psychology of Music*, *21*, 103-113.

DAVIDSON, J. W. (1994). What type of information is conveyed in the body movements of solo musician performers? *Journal of Human Movement Studies*, *6*, 279-301.

DAVIDSON, J. W. (1995). What does the visual information contained in music performances offer the observer? Some preliminary thoughts. In R. Steinberg (Ed.), *Music and the mind machine: Psychophysiology and psychopathology of the sense of music* (pp. 105-114). Heidelberg: Springer.

DAVIDSON, J. W., & CORREIA, J. S. (2002). Body movement. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance. Creative strategies for teaching and learning* (pp. 237-250). Oxford: Oxford University Press.

DE MEIJER, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, *13*, 247-268.

DE MEIJER, M. (1991). The attritution of aggression and grief to body movements: The effects of sex-stereotypes. *European Journal of Social Psychology*, *21*, 249-259.

DITTRICH, W. H., TROSCIANKO, T., LEA, S. E., & MORGAN, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *6*, 727-738.

FISHBEIN, M., MIDDLESTADT, S., OTTATI, V., STRAUS, S., & ELLIS, A. (1988). Medical problems among ICSOM musicians: Overview of a national survey. *Medical Problems of Performing Artists*, *3*, 1-8.

FRIBERG, A. (2005). A fuzzy analyzer of emotional expression in music performance and body motion. In J. Sundberg & B. Brunson (Eds.), *Proceedings of Music and Music Science* [CD]. Stockholm: Royal College of Music.

FRIBERG, A., & SUNDBERG, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustical Society of America*, *105*, 1469-1484.

GABRIELSSON, A., & JUSLIN, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*, 68-91.

GABRIELSSON, A., & LINDSTRÖM, E. (2001). The influence of musical structure on emotional expression. In P. Juslin & J. A. Sloboda (Eds.), *Music and emotion* (pp. 223-248). Oxford: Oxford University Press.

GOLDENBERG, MORRIS (1950). *Modern school for xylophone, marimba, vibraphone*. New York : Chappell.

JOHANSSON, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, *14*, 201-211.

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 theoretical framework. In P. Juslin & J. A. Sloboda (Eds.), *Music and emotion* (pp. 309-337). Oxford: Oxford University Press.

JUSLIN, P. N., FRIBERG, A., & BRESIN, R. (2002). Toward a computational model of expression in performance: The GERM model. *Musicae Scientiae*, *special issue 2001-2002*, 63-122.

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., & SLOBODA, J. A. (EDS.). (2001). *Music and emotion*. Oxford: Oxford University Press.

LEE A. (2006). VirtualDub: A desktop video processing and capture application [Computer software]. Retrieved from http://www.virtualdub.org/

MACFARLANE, L., KULKA, I., & POLLICK, F. E. (2004). The representation of affect revealed by butoh dance. *Psychologia*, *47*, 96-103.

MCNEILL, D., QUEK, F., MCCULLOUGH, K.-E., DUNCAN, S., BRYLL, R., MA, X.-F., & ANSARI, R. (2002). Dynamic imagery in speech and gesture. In B. Granström, D. House, & I. Karlsson (Eds.), *Multimodality in language and speech systems* (pp. 27-44). Boston: Kluwer Academic Publishers.

PATERSON, H. M., POLLICK, F. E., & SANFORD, A. J. (2001). The role of velocity in affect discrimination. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the cognitive science society*, Edinburgh (pp. 756-761). Mahwah, N.J. : Lawrence Erlbaum Associates.

POLLICK, F. E., PATERSON, H. M., BRUDERLIN, A., & SANFORD, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*, B51-B61.

R Development Core Team. (2006). R: A language and environment for statistical computing [Computer software]. Retrieved from http://www.R-project.org/

Repp, B. (1993). Music as motion: a synopsis of Alexander Truslit's (1938) "Gestaltung und bewegung in der musik". *Psychology of Music*, *21*, 48-72.

Resnicow, J. E., Salovey, P., & Repp, B. H. (2004). Is recognition of emotion in music performance an aspect of emotional intelligence? *Music Perception*, *22*, 145-158.

Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology*, *7*, 733-740.

Saldaña, H. M., & Rosenbaum, D., Lawrence. (1993). Visual influences on auditory pluck and bow judgements. *Perception and Psychophysics*, *54*, 406-416.

Schutz, M., & Lipscomb, S. (in press). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception.*

Shinosako, H., & Ohgushi, K. (1996). Interaction between auditory and visual processing in impressional evaluation of a piano performance. In S. Kuwano & T. Kato, *Proceedings of the Third Joint Acoustical Society of America and the Acoustical Society of Japan* (pp. 357-361). Tokyo: Acoustical Society of Japan

Shove, P., & Repp, B. (1995). Musical motion and performance: theoretical and empirical perspectives. In J. Rink (Ed.), *The practice of performance. Studies in musical interpretation* (pp. 55-83). Cambridge: Cambridge University Press.

Sörgjerd, M. (2000). *Auditory and visual recognition of emotional expression in performances of music*. Unpublished master's thesis, Uppsala University, Uppsala, Sweden.

Stoffregen, T. A., Pagulayan, R. J., Bardy, B. G., & Hettinger, L. J. (2000). Modulating postural control to facilitate visual performance. *Human Movement Science*, *19*, 203-220.

Sundberg, J., Iwarsson, J., & Hagegård, H. (1995). A singer's expression of emotions in sung performance. In O. Fujimura & M. Hirano (Eds.), *Vocal fold physiology: Voice quality control* (pp. 217-229). San Diego: Singular Press.

Todd, N. P. M. (1999). Motion in music: A neurobiological perspective. *Music Perception*, *17*, 115-126.

Walk, R. D., & Homan, C. P. (1984). Emotion and dance in dynamic light displays. *Bulletin of Psychonomic Society*, *22*, 437-440.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. In I. Wachsmuth & T. Sowa (Eds.), *Gesture and sign language in human-computer interaction* (pp. 241-253). Berlin, Heidelberg: Springer Verlag.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, *34*, 97-113.