

Visual perception of expressiveness in musicians' body movements.

Sofia Dahl and Anders Friberg
KTH School of Computer Science and Communication
Dept. of Speech, Music and Hearing
Royal Institute of Technology
Sweden

Musicians often make gestures and move their bodies expressing the musical intention. This visual information provides a channel of communication to the listener of its own, separated from the auditory signal. In order to explore to what extent emotional intentions can be conveyed through musicians' movements, subjects watched and rated silent video clips of musicians performing four different emotional intentions, Happy, Sad, Angry, and Fearful. In a first experiment twenty subjects were asked to rate performances on the marimba with respect to perceived emotional content and movement character. Video clips were presented in different *viewing conditions*, showing selected parts of the player. The results showed that the intentions Happiness, Sadness and Anger were well communicated, while Fear was not. The identification of the intended emotion was only slightly influenced by viewing condition. The movement ratings indicated that observers used cues to distinguish between intentions. In a second experiment subjects rated the same emotional intentions and movement character for woodwind performances, one bassoon player and one soprano saxophonist. The ratings from the second experiment confirmed that Fear was not communicated while Happiness, Sadness and Anger were recognized. The movement cues used by the subjects in the first experiment appeared also in the second experiment and were similar to cues in audio signals conveying emotions in music performance.

Introduction

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, subjects 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), and landing positions of strokes of badminton playing (Abernethy & Russel, 1987). It is also possible to identify the emotional expression in dance and music performances (Walk & Homan, 1984, Dittrich, Troscianko, Lea, & Morgan, 1996, Sörgjerd, 2000), as well as the emotional expression in every-day arm movements such as drinking and lifting (Pollick, Paterson, Bruderlin, & Sanford, 2001, Paterson, Pollick, & 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 during a tone on a violin has a direct relationship to 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 speech. In this respect the movements and the spoken words are co-expressive, not subordinate to each other. Bearing in mind that music also is a form of com-

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 II, and Anders Askenfelt and Peta Sjölander for valuable comments on the manuscript.

This work was supported by the European Union (MEGA - Multisensory Expressive Gesture Applications, IST-1999-20410). The collaboration with Andrea Bressan was made possible through the SOCRATES programme: Higher Education (ERASMUS).

For editing the video clips used in the experiments, the open source software Virtual Dub was used (<http://www.virtualdub.org/>). For parts of the analysis the open source statistical software R was used (<http://www.R-project.org/>).

munication 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 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. However, 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 made several studies on expressive movements in musical performance relating the overall perceived expressiveness to musicians' movements (e.g. Davidson, 1993, 1994, Clarke & Davidson, 1998). Most of these studies used video recordings, utilizing the point-light technique (Johansson, 1973) to capture the movements of musicians (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. Subjects rated these performances on a scale of expressiveness (ranging from "inexpressive" to "highly expressive"). From this data Davidson (1993) concluded that subjects 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 naive subjects even performed better when only watching, compared to the other conditions, 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, 1995).

Davidson (1994) also investigated which parts of a pianist's body that conveyed the information the observers 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 subjects. One clarinet player and one violinist performed pieces with the emotional intentions Happiness, Sadness, Anger, Fear, Solemnity, Tenderness, and No expression. Subjects were asked to select the most appropri-

ate emotion for each performance. Sörgjerd found that subjects 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 less strong for Angry and Fearful music.

De Meijer and Boone and Cunningham (1999) proposed several movement cues considered important for detecting emotional expression (De Meijer, 1989, 1991, Boone & Cunningham, 1999, 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 match well the findings by De Meijer, concerning viewers' attribution of emotion to specific body movements (1989, 1991). 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 perspective of the previously mentioned studies using musicians' movements. The arm move-

ments 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, while 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 is a natural choice, as indicated by performance instructions such as *andante* (walking), or *corrente* (running). Also 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 performer or what part of the player the observers see?
3. How can perceived emotions be described in terms of movement cues?

Two experiments were performed to answer these questions. In Experiment I subjects rated performances on marimba, and in Experiment II subjects rated woodwind performances.

Experiment I

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 subjects were presented with video clips showing the player to different extent.

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 the expression in the sounding performance rather than in body movements. The player was aware that the performances would be filmed 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". 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 made the camera ready 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 a freeware video editing software (Virtual Dub). 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 x 2 repetitions x 4 conditions) video clips were generated. The duration of the video clips varied between 30 and 50 s.

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

Procedure. Subjects 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.

The subjects were also asked to rate the perceived movement character. Four movement cues were selected, taking into account that (a) they 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.



original



full



nohands



torso



head

Figure 1. Original (top) and filtered video images exemplifying the four viewing conditions used in the test: full, nohands, torso, and head.

The cues were, with their musical counterpart in parenthesis; Amount (sound level), Speed (tempo), Fluency (articulation), Regularity (tempo variations). The ratings of the cues were carried out using bipolar scales, coded from 0 to 6:

Amount: none - large
 Speed: slow - fast
 Fluency: jerky - smooth
 Regularity: irregular - regular

The assumption was 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 variation in movement patterns over the performance.

The 32 video clips were presented on a computer screen and rated individually. For each subject a command-file automatically opened the clips in the Windows mediaplayer in a randomized order. Each clip could be viewed as many times as the subject liked, but once the window for a specific clip had been closed, the next clip started automatically and the subject could no longer go back to rate the previous one.

Measure of achievement.

The use of rating adjectives on individual scales results in many values for each stimulus, presentation and subject. In order to make a clear overview of data such as this, with several factors involved, it may be useful to calculate individual measures of how well the communication succeeded in each case. One benefit with such a measure is that it is easy to investigate all the independent factors in one analysis of variance, summarizing all rated adjectives or emotions.

Previous examples of how to combine several rated scales can 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.

One drawback with these estimations is that they do not consider the absolute magnitude of the ratings as will be shown below. Instead, we suggest to use 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, r_{xy} being 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.

Table 1

Comparison between achievement (A), point-biserial correlation (r , used by Juslin, 2000), and the emotion recognition score (E , used by Resnicov, Salovey & Repp, 2004) calculated for combinations of the intention vector for Anger $\mathbf{x} = [F A H S] = [0 1 0 0]$ and different rating vectors. While Cov reflects differences in magnitude in ratings, r and E will generate the same value for different cases.

$\mathbf{y} =$	$[F A H S]$	A	r	E
intention	0 6 0 0	1.00	1.00	1.00
correctly	0 1 0 0	0.17	1.00	1.00
identified	0 3 0 0	0.50	1.00	1.00
(ranked highest)	2 3 2 2	0.17	1.00	0.33
ambiguous	0 6 6 0	0.67	0.58	0.50
or	0 3 3 0	0.33	0.58	0.50
equal ranking	1 1 1 1	0.00	-	0.25
confusion	1 0 0 0	-0.05	-0.33	0.00
or non-	3 0 0 0	-0.17	-0.33	0.00
successful	6 0 0 0	-0.33	-0.33	0.00
communication	6 5 5 5	-0.05	-0.33	0.24

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

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

where \mathbf{x} and \mathbf{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 \mathbf{x} can only take the values 0 and 1, and \mathbf{y} can be integer values between 0 and 6, $C = 1.5$ in all cases.

A comparison between values for the covariance and correlation of different vectors is shown in Table 1. The table shows A , the correlation coefficient r , and the emotion recognition score E between the intention vector for anger $\mathbf{x} = [F A H S] = [0 1 0 0]$ and different \mathbf{y} . As seen in the table, Cov reflects the magnitude of the ratings. A rating of $\mathbf{y} = [0 6 0 0]$ gives a higher Cov value than $\mathbf{y} = [0 3 0 0]$, while 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. Cov , 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 $\mathbf{y} = [0 6 6 0]$ and $\mathbf{y} = [0 3 3 0]$). Note that an equal rating of all four emotions,

e.g. a rating vector $\mathbf{y} = [1 \ 1 \ 1 \ 1]$, (with a standard deviation of 0), does not yield any numerical value for r . Therefore r makes a less strong candidate for a measure. The emotion recognition score E always yields numerical values, but also 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 equal. We assume that a achievement significantly larger than zero 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 do not take any negative values for confusing cases, A is a more strict measure.

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 subjects and the two performances of each intended emotion. The 95 % confidence intervals are indicated by the vertical error bars. The figure illustrates that the player was able to convey three of the four intended emotions to the subjects in most viewing conditions. Sadness was most successfully identified, 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.

To investigate the effects of the intended emotions and viewing conditions, the achievement measures were subjected to a 4 conditions \times 4 emotions repeated measures ANOVA. The analysis showed main effects for intended emotion [$F(3, 36) = 19.05, p < 0.0001$] and viewing conditions [$F(3, 36) = 6.98, p < 0.001$], and significant results for the two-way interaction viewing condition \times emotion [$F(9, 108) = 4.36, p < 0.0001$].

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 ($p < 0.0001$) achievement than all the other three intentions.

The interpretation of the effect of viewing condition was somewhat complicated. A Tukey post hoc test showed that the torso and head conditions received significantly lower achievement compared to the full condition ($p < 0.0001$). No other differences between viewing conditions were significant. Thus this confirmed the a priori assumption that seeing more of the body of the performer improves the achievement.

The interaction between emotion and viewing condition is illustrated in Table 2. The table shows the mean achievement, averaged across 20 subjects and two performances, for each intended emotion and viewing condition. The significant 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)

Table 2

Mean achievement for the four intended emotions and viewing conditions (full, nohands, torso, and head) averaged across 20 subjects and two performances. The viewing condition receiving the highest achievement for each specific intention is shown in bold. Seeing the full view of the player did not automatically result in high achievement. 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.

Intent	full	nohands	torso	head	mean
Happiness	.46	.32	.48	.35	.40
Sadness	.57	.64	.34	.65	.55
Anger	.57	.44	.27	.29	.40
Fear	.15	.08	.07	-.04	.07
column mean	.44	.37	.29	.31	

Table 3

Effect sizes d for the four viewing conditions. The table shows the differences between the mean achievement for specific viewing conditions. The values are rather small, meaning that the overlap between the distributions for different viewing conditions was large.

	full	nohands	torso	head
full	-			
nohands	.18	-		
torso	.40	.19	-	
head	.32	.14	.04	-

received high achievement values (from 0.57 to 0.65 in Table 2), while the mean achievement for the torso condition was much lower (0.34). A Tukey post hoc test revealed that the only significant effect within the Sad intention was between torso and head ($p < 0.05$). For the Happy intention, on the other hand, the torso received higher achievement (.48) than the full condition (.46), however no post hoc test were significant. For Anger, the full condition received the highest achievement (0.57), while the torso and head conditions were less successful in conveying the intention. The only post hoc effect was between torso and full condition ($p < 0.05$).

The results for viewing condition were somewhat surprising. Initially one would hypothesize that seeing more of the player would provide the subjects 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, looking at the interactions, Anger was the only intention showing a similar relationship between viewing conditions (see Table 2 and Figure 2).

Table 3 display the effect sizes, d , calculated as the difference in mean achievement between two viewing conditions, divided by the pooled standard deviations. Following the

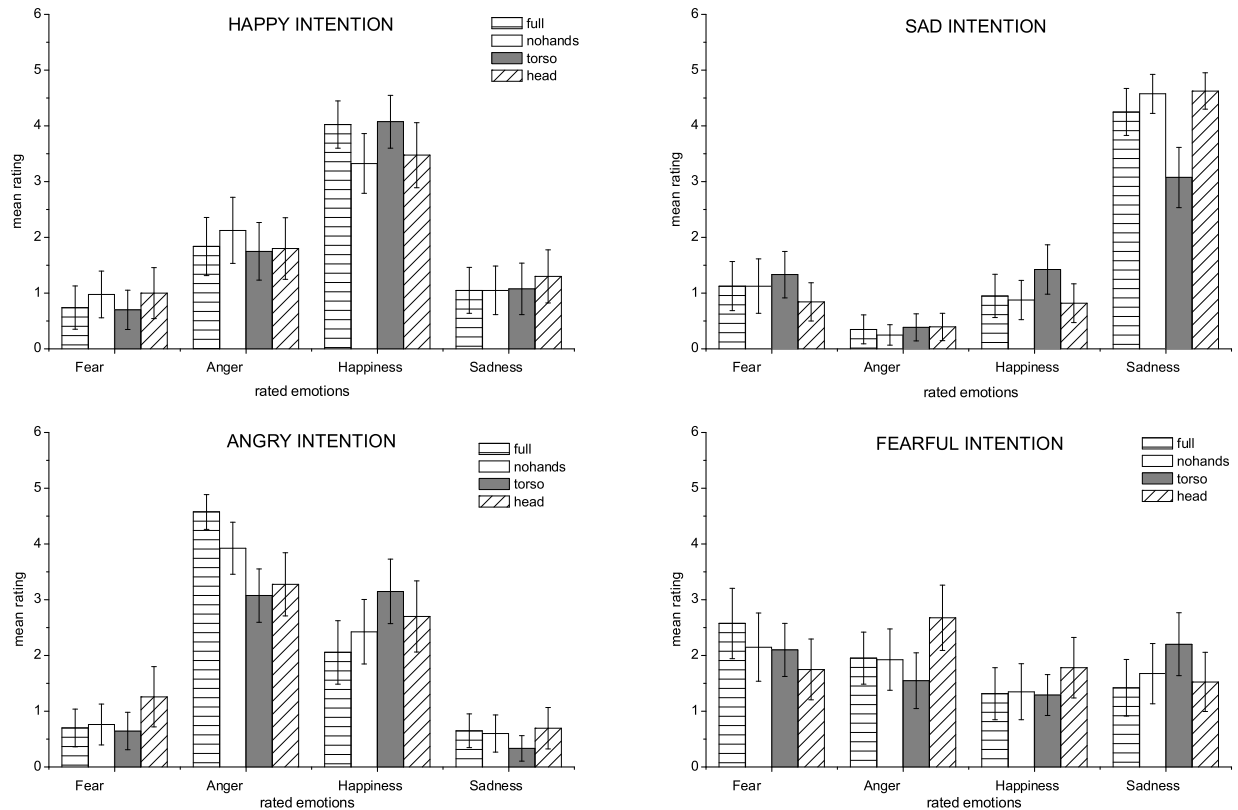


Figure 2. Ratings for the four intended emotions and viewing conditions. Each panel shows the mean ratings for the four emotions averaged across 20 subjects and the two performances of each intended emotion. The patterns of the bars show the four viewing conditions: full (horizontally striped), nohands (white), torso (grey), and head (diagonally striped). The error bars indicate 95 % confidence interval. As seen in the panels the Happy (top left panel), Sad (top right) and Angry (bottom left) performances receive ratings in correspondence with the intention, while Fearful (bottom right) was hardly recognized at all.

classifications of Cohen (1988), the effect sizes were small, meaning that the overlap between distributions was at least half a standard deviation.

Also effect sizes for the interaction viewing condition and intended emotion displayed small to medium values for d in most cases. The exceptions were the intentions Sadness and Anger. For Sadness, there were medium to large differences between torso and each of the three viewing conditions where the head was visible; full ($d=.7$), nohands ($d=.98$), and head ($d=1.00$). For Anger, the large differences appeared between the conditions full and torso ($d = 1.00$), or between the conditions full and head ($d=.88$).

Looking at the results in an alternative way, the subjects' ratings were transformed into forced choice responses, commonly used in other studies. The transformation was done in a strict fashion, meaning that only the ratings where the intended emotion received the highest rating was considered as "correct". The percentages of correct responses are shown in Table 4. The pattern of these values corresponds very closely to the mean achievement across the performances seen in Ta-

Table 4

Correct identification of the intended emotions in percent for the four viewing conditions, averaged across the two performances of each intention. The values were calculated as the portion of ratings where the intended emotion received the highest rating. The viewing condition receiving most correct identifications for a specific intention is shown in bold.

	full	nohands	torso	head	row mean
Happiness	68	50	73	56	61.8
Sadness	80	80	53	95	77.0
Anger	85	60	38	45	57.0
Fear	35	23	23	10	22.8
column mean	67.0	53.3	46.8	51.5	

ble 2. Sadness, Anger, and Happiness were identified well above chance level (25%).

Movement cues. Figure 3 shows the mean ratings of the movement cues for each intended emotion. The movement

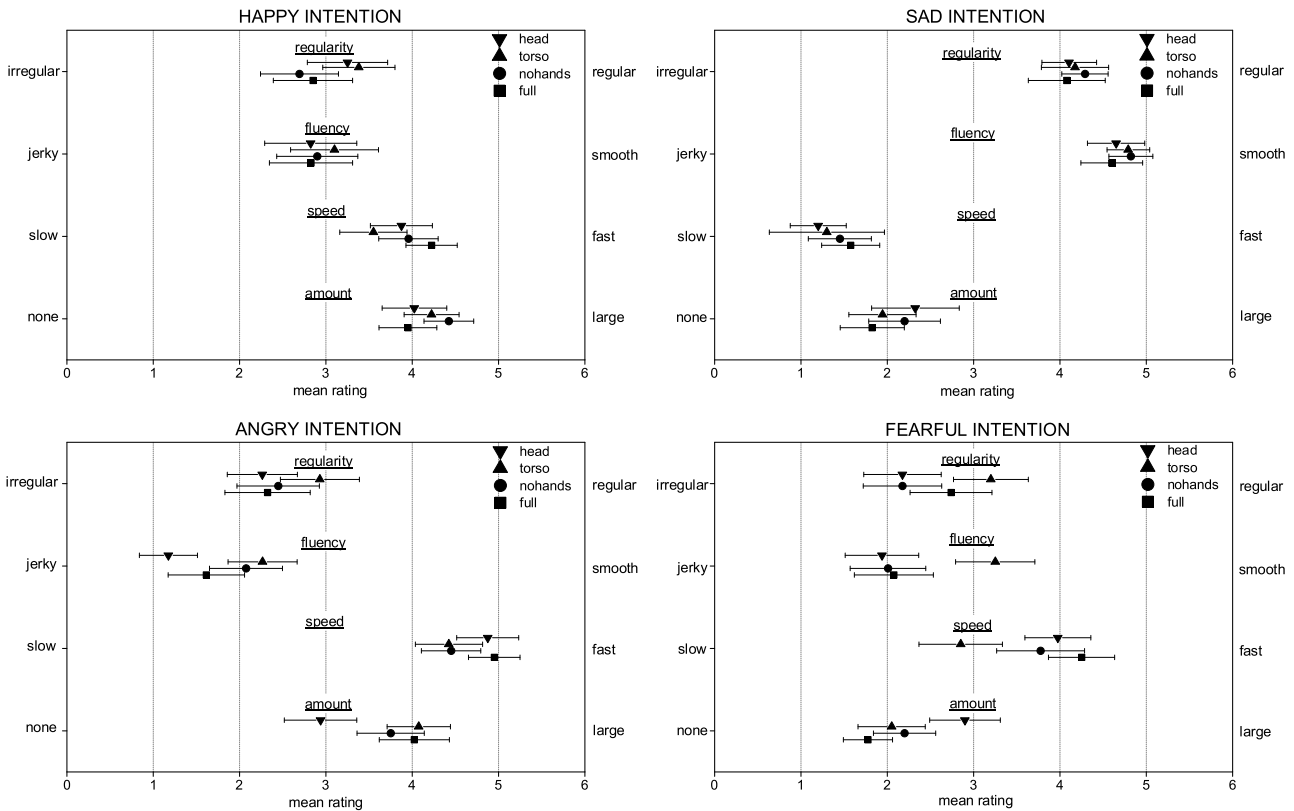


Figure 3. Ratings of movement cues for each intended emotion and viewing condition. Each panel shows the mean markings for the four emotions averaged across 20 subjects and the two performances of each intended emotion. The four viewing conditions are indicated by the symbols: full (square), nohands (circle), torso (pyramid), and head (top-down triangle). The error bars indicate 95 % confidence interval. As seen in the panels, the movement characterization differs for the four intentions.

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 correspond to large amount of movement, high speed, smooth fluency, and regular movements, while low ratings correspond 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 somewhat 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 5 shows the intercorrelations between the movement cues. As expected, they were all somewhat correlated with values ranging from -.62 to .26. The amount of movement seems to be relatively independent, reflected in the small correlations with the other cues. Speed, Fluency and Regularity

Table 5
Intercorrelations between the movement cues rated in Experiment 1. All correlations were statistically significant ($p < 0.01, N = 618$)

	amount	speed	fluency	regularity
amount	-			
speed	.26**	-		
fluency	-.19**	-.62**	-	
regularity	-.12**	.44**	.58**	-

** $p < 0.01$

all show relatively medium intercorrelations.

In order to investigate how the rated emotions were related to the rated movement cues, a multiple regression analysis was performed. Each rated emotion was predicted using the four movement ratings as independent variables. In Table 6 the resulting multiple correlation coefficients (R), the standardized beta-weights, and the semipartial correlations are presented for each emotion.

The overall multiple correlation coefficients yielded rather

Table 6

Results from the regression analysis for the rated emotions and rated movement cues in Experiment I. The numbers display beta-weights and the semipartial correlations (in italics) of the regression analysis.

	amount	speed	fluency	regularity
Happiness <i>R</i> = 0.44	.35*** <i>.34***</i>	.21*** <i>.16***</i>	.06 <i>.04</i>	-.01 <i>-.01</i>
Sadness <i>R</i> = 0.65	-.18*** <i>-.18***</i>	-.43*** <i>-.33***</i>	.15*** <i>.11***</i>	.08* <i>.06*</i>
Anger <i>R</i> = 0.61	.19*** <i>.18***</i>	.18*** <i>.14***</i>	-.30*** <i>-.21***</i>	-.16*** <i>-.13***</i>
Fear <i>R</i> = 0.32	-.29*** <i>-.28***</i>	-.05 <i>-.04</i>	-.17** <i>-.12**</i>	-.06 <i>-.05</i>

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

low values in terms of explained variance, ranging from 10 to 42%. Applying multiple correlation on averaged ratings over subjects 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 6). 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 in 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 II

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 I 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, three saxophonists were recorded. However, two of these barely moved at all and were not used in the following experiment.

Four melodies were used for the performances: Berwald's String Quartet No. 5, C major, bars 58 to 69; Brahms' Symphony Op. 90 No. 3 in C minor, first theme of the third movement, *Poco allegretto*; Haydn's Quartet in F major for strings, Op. 74 No. 2, theme from first movement, and Mozart's sonata for piano in A major, K331, first eight bars. Unlike the piece used for the performances in Experiment I, 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 (c.f. 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). Among the 12 intentions were the four used in Experiment I; Happiness, Sadness, Anger, and Fear. 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. Also an "indifferent" performance was recorded.

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

The movements were recorded using the same digital video camera as in Experiment I. 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 was used.

From the 12 emotional expressions recorded, the performances of Happiness, Sadness, Anger and Fear were selected as video stimuli. The editing of the video clips was similar to that in Experiment I. 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 x 2 players x 4 excerpts) video clips were generated. The duration of the video clips varied between 9 and 46 s.



saxophone original



saxophone filtered



bassoon original



bassoon filtered

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

Subjects. A total of 20 (10 male and 10 female) subjects volunteered to participate in the experiment. The subjects were between 23 and 59 years old (mean 31, standard deviation 8) with varying amounts of musical training. Five subjects reported that they did not play any instrument. Ten of the subject had several years of experience with one or more instrument and played regularly. Subjects recruited from out-

side the department received a small compensation for their participation. None of the subjects had participated in Experiment I.

Procedure. The subjects were asked to rate the same parameters as those in Experiment I. 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 was 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 I, 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 rated individually. The stimuli clips were presented in two blocks, one block with all saxophone clips, randomized for each subject and session, and another block with all bassoon clips. Half of the subjects started rating the block with saxophone clips first and the remaining half started by first rating the bassoon clips. Each clip was repeatedly played until the subject 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 subject could no longer go back to rate the previous one. In order to get the subjects acquainted to the procedure, a pre-test was run. During the pre-test the subject was able 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 two players and four emotions, averaged across the 20 subjects 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 subjects, while Fear was not. In general, however, the ratings were lower compared to Experiment I. There seem 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 I, the achievement was calculated for each of the presented video clips (see Equation 1). The achievement measure was then subjected to an analysis of variance to investigate the effects of the intended emotions and musical excerpts. The 4 excerpts \times 4 emotions \times 2 performers repeated measures ANOVA showed main effects for intended emotion [$F(3,57) = 42.06, p < 0.00001$], musical excerpt [$F(3,57) = 11.53, p < 0.00001$], and performer [$F(1,19) = 5.45, p < 0.05$], and significant results for all two-way interactions: excerpt \times emotion [$F(9,171) = 6.65, p < 0.00001$], emotion \times performer [$F(3,57) = 12.61, p < 0.00001$], and excerpt \times performer [$F(3,57) = 7.21, p < 0.0005$].

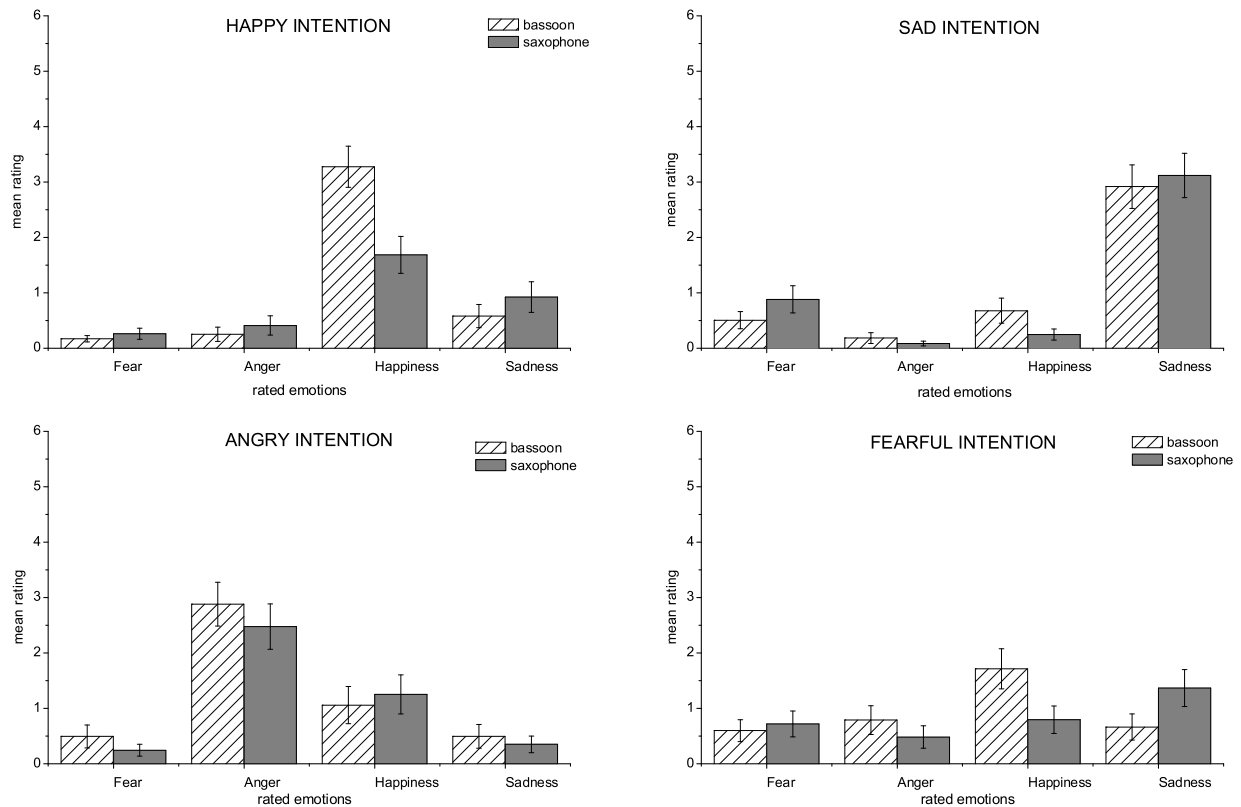


Figure 5. Ratings for the four intended emotions and two instrumentalists; bassoon player (striped bars) and saxophonist (grey bars). Each panel shows the mean ratings for the four emotions averaged across 20 subjects and the four musical excerpts of each intended emotion. The error bars indicate 95 % confidence interval. As seen in the panels the Happy (top left panel), Sad (top right) and Angry (bottom left) performances receive ratings in correspondence with the intention, while subjects failed to recognize the Fearful intention (bottom right). In general, subjects rated the bassoonist as slightly more Happy, while the saxophonist received higher ratings for Sad.

Table 7

Mean achievement for the two performers, excerpts and intended emotion. The table shows the achievement for the bassoon player (left half) and saxophonist (right half) and the musical excerpts Berwald, Brahms, Haydn, and Mozart, averaged across 20 subjects. The performance receiving the highest achievement for each specific intention is shown in bold.

intent	bassoon				saxophone				row mean
	Berwald	Brahms	Haydn	Mozart	Berwald	Brahms	Haydn	Mozart	
Happiness	.69	.44	.63	.22	.17	.15	.33	.12	.34
Sadness	.33	.39	.48	.44	.47	.30	.44	.60	.43
Anger	.44	.16	.48	.39	.24	.12	.48	.41	.34
Fear	-.04	-.03	-.07	-.17	.00	-.13	.02	.01	-.05
column mean	.35	.24	.38	.22	.22	.11	.31	.29	

Similarly to Experiment I, 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 ($p < 0.0005$) compared to those of the other three intentions.

In Table 7 the achievement values for all factors are shown. The table shows the mean achievement values for the two performers, four excerpts and each intended emotion.

The mean achievement for the Fearful intention (-.05) is considerably lower compared to the intentions Happiness (.34), Sadness (.43), and Anger (.34).

The main effect of performer was due to slightly higher achievement values for the bassoon player compared to the saxophonist, see Table 7. 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 (c.f. Figure 5). The Happy intention for the saxophonist received significantly lower achievement values than all other cases ($p > 0.001$), 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 the Brahms excerpt received significantly lower achievement compared to the other excerpts ($p < 0.05$). In addition, the Haydn excerpt received significantly higher achievement values than the Mozart excerpt ($p < 0.05$).

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 ($p < 0.001$). For the Brahms excerpt Anger received significantly lower achievement than the Sad intention ($p < 0.01$). 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.

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 I, especially for Anger and Sadness (c.f. Figure 3).

Table 8 shows the intercorrelations between the rated movement cues for the two performers in Experiment II. For the saxophonist, the movement cues are all somewhat correlated with values ranging from -0.80 to 0.58. For the bassoon player three of the cues, speed, fluency, and regularity, are correlated with values ranging from -0.77 to 0.20. Amount, however, is independent from the other cues (similar to Experiment I, see Table 5), suggesting differing movement characteristics between the two players.

As in Experiment I, the relation between rated emotion and movement cues was investigated in a multiple regression analysis. Table 9 displays the resulting multiple correlation coefficients, the standardized beta-weights, and the semipartial correlations (sr) for the two performers in Experiment II. In general, the results for the bassoon player and the saxophonist are similar to those for the marimba player in Experiment I. The explained variance was also similar to Experiment I, 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 Anger is Fluency (jerky $sr = -.18$ and $-.25$). Neither Amount, nor Speed, was significantly contributing to Anger ratings. For Happiness a difference between the two per-

Table 8

Intercorrelations between the movement cues for the two performers rated in Experiment II, bassoon players (top) and saxophonist (bottom). Half of the correlations for the bassoon player and all correlations for the saxophonist were statistically significant ($p < 0.01, N = 320$)

bassoon	amount	speed	fluency	regularity
amount	-			
speed	-.01	-		
fluency	.02	-.77**	-	
regularity	-.07	-.14**	.20**	-
saxophone	amount	speed	fluency	regularity
amount	-			
speed	.58**	-		
fluency	-.43**	-.80**	-	
regularity	-.18**	-.28**	.41**	-

** $p < 0.01$

Table 9

Results from the regression analysis for the rated emotions and rated movement cues in Experiment II. The numbers display beta-weights and the semipartial correlations (in italics) of the regression analysis for the bassoon player (top) and saxophonist (bottom).

	movement cues			
	amount	speed	fluency	regularity
bassoon				
Happiness	.04	.36***	-.03	.01
$R = 0.38$.04	.23***	-.02	.01
Sadness	-.02	-.49***	.24***	.00
$R = 0.69$	-.02	-.31***	.15***	.00
Anger	.06	.14	-.29***	-.08
$R = 0.43$.06	.09	-.18***	-.08
Fear	-.12*	-.12	.04	-.02
$R = 0.19$	-.12*	-.08	.02	-.02
saxophone				
Happiness	.24***	.19	.05	-.01
$R = 0.35$.20***	.10	.03	-.01
Sadness	-.08	-.37***	.29***	-.09
$R = 0.65$	-.07	-.20***	.16***	-.08
Anger	.01	.17	-.44***	.00
$R = 0.59$.01	.09	-.25***	.00
Fear	-.23***	.10	.24*	-.08
$R = 0.30$	-.18***	.05	.13*	-.07

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

formers' movement cues can be seen. For the bassoon player, the most important cue to predict Happiness was Speed (fast $sr = .23$), while 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

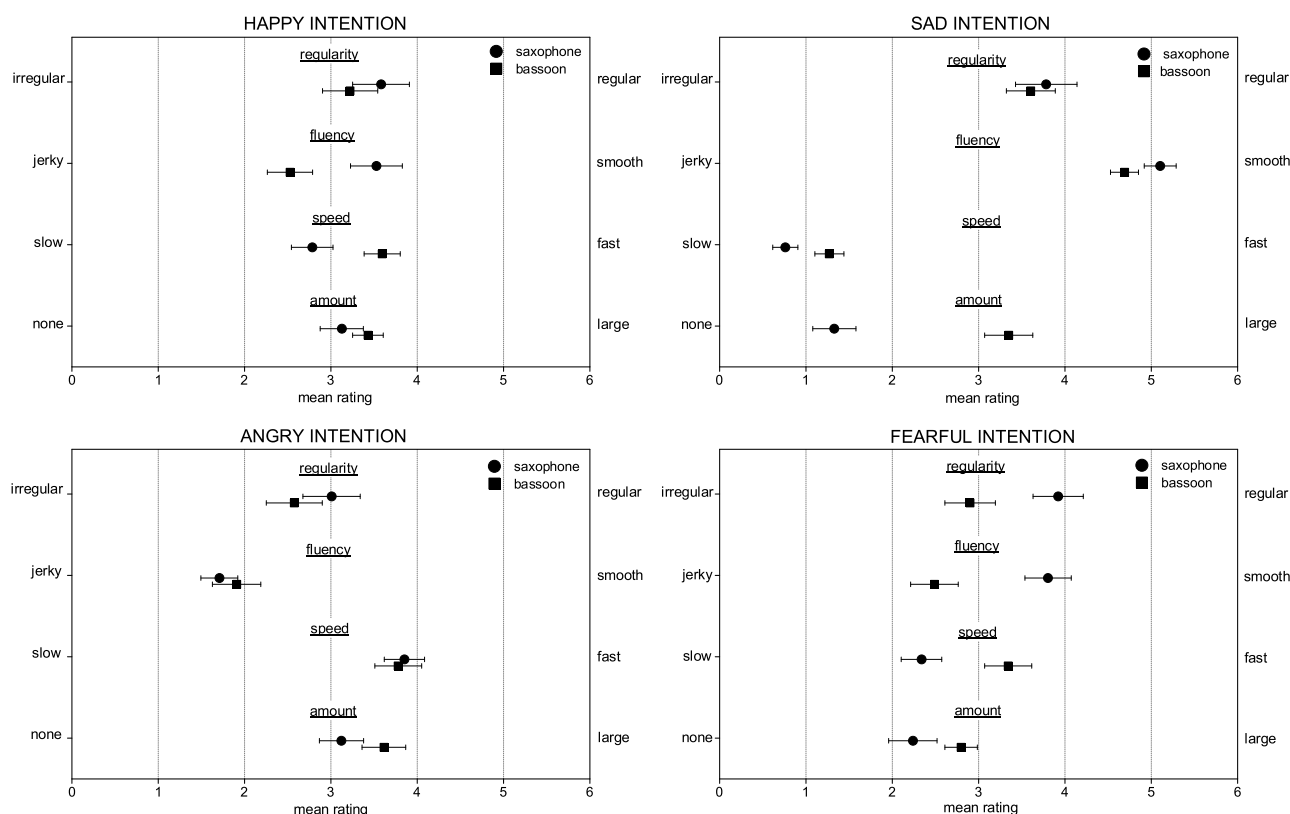


Figure 6. Mean ratings for the movement cues and two performers in Experiment II, averaged across four excerpts and 20 subjects. The error bars indicate 95 % confidence interval. The ratings are similar to those for the marimba player (see Figure 3). In general, the movements for the bassoon player (squares) and the saxophonist (filled circles) were rated similarly.

the very low ratings for Fear, however, its characterization is questionable.

Results across performers

The combined result of Experiment I and II can be seen in Figure 7. The figure shows mean ratings for the four rated emotions averaged across all players, subjects and cases. As seen in the figure, the intentions Happiness, Sadness, and Anger were identified while Fear was not recognized at all. Transformed into forced choice responses, the total percentage of correct identifications of the four intended emotions, across both experiments, were 64% for Happiness, 78% for Sadness, 62% for Anger, and 19% for Fear (see Table 10). The performer who received most correct identified intentions varied with emotion. The bassoon player received the highest values for Happiness (81%) and Anger (72%), for Sadness the saxophonist received 83% correct, but recognition was high for all three performers. The best recognition for Fear, 22% correct for the marimba performances, was still below chance level (25%).

Table 11 presents the result of a multiple regression analysis of all the rated movement cues and emotional intentions.

Table 10

Correct identification of the intended emotions in percent for each of the three performers rated in Experiment I and II. The values were calculated as the portion of ratings where the intended emotion received the highest rating. The performer receiving most correct identifications for a specific intention is shown in bold.

intent	marimba	bassoon	saxophone	row mean
Happiness	60	81	56	64.4
Sadness	76	80	83	78.4
Anger	56	72	64	61.9
Fear	22	14	19	19.1
column mean	53.3	61.9	55.3	

Comparing the beta-weights and semipartial correlations in Table 11 to the individual results in Experiment I and II there are many similarities. Sadness is characterized by slow and smooth movements, while Anger is characterized by jerky movements. Large, and to a lesser degree, somewhat fast and jerky movements characterize Happiness. Most important to characterize Fear are small movements, but uneven regularity

Table 11

Comparison between the movement cues for Experiment I and II, and reported auditory cues expressing emotional intentions. The numbers display beta-weights and the semipartial correlations (in italics) of the regression analysis. For comparison, the corresponding audio cues; sound level, tempo, articulation, and tempo variability (selected from Juslin, 2001), are displayed to the right in the table.

	movement cues				auditory cues			
	amount	speed	fluency	regularity	sound lev.	tempo	articul.	tempo var.
Happiness N=315 <i>R</i> = 0.34	.22*** .22***	.13 * .12*	-.15* -.13*	.06 .01	high	fast	staccato	small
Sadness N=314 <i>R</i> = 0.36	-.02 -.02	-.22*** -.20***	.19** .16**	-.01 -.01	low	slow	legato	final ritard
Anger N=312 <i>R</i> = 0.30	.10 .10	.05 .04	-.25*** -.21***	-.06 -.05	high	fast	staccato	small
Fear N=317 <i>R</i> = 0.26	-.24*** -.23***	.04 .03	.01 .01	-.14* -.12*	low	fast	staccato	large

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

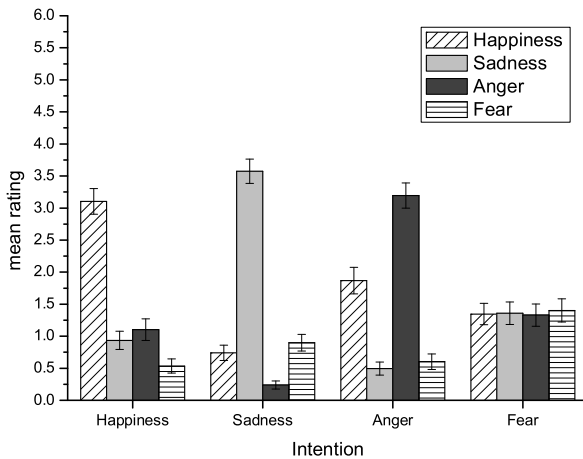


Figure 7. Mean ratings for the four rated emotions averaged across all players, subjects and cases in Experiment I and II. Each bar is based on 320 values. The error bars indicate 95 % confidence interval.

also has 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 11: 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 (as expressed by the *sr* values) and their corresponding audio counterparts all change in the same direction.

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 performer/instrument and what parts of the player the subjects see, and 3) if movement cues can be used to describe this communication. Our results indicate that

1. the intended emotions Sadness, Happiness, and Anger were successfully conveyed, while 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.

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, subjects have identified the intentions well above chance in many cases (e.g. Walk & Homan, 1984, Dittrich et al., 1996, Boone & Cunningham, 1998, 2001). The emotion that has been most correctly identified differs between studies, but Anger, Sadness/Grief and Happiness/Joy generally receive 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 being 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 which allow adult observers identify Sadness at above chance level (Boone & Cunningham, 2001). The ability to identify and portray Fear, Anger and Happiness appears later, from the age of five.

The confusion between Happiness and Anger seen in our results has also been found by other researchers. 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 Sörgerd (2000) found that Anger and Happiness was often confused with each 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. Sörgerd (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. What could be the reason that the communication of Fear fail in the present study?

Music performances are capable of expressing a wide range of emotional expressions. Sounding performances can communicate both expressive intentions inherent in the original score as well as those added during the performance (Juslin & Sloboda, 2001). Some music is even composed with the explicit intention of *inducing* Fear (e.g. Herrmann's original film music to "Psycho"). However, 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. Other intense emotions, such as Anger, Love, Happiness, Sadness and Grief all make sense in the perspective of non-verbal communication through music performance. But who expects someone who is truly frightened to stay on stage, pick up an instrument and play? In fact, we are not likely to consider performances that show signs of anxiety or fear to be good or professional.¹ 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 II. In the pilot test subjects rated all 12 recorded intentions. From the result of the pilot test it was evident that subjects were more likely to consider positive emotions such as Love, Pride and Contentment as expressed, while 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 strong correspondence was found between how the selected movement cues and their audio counterparts varied for different intentions (see Table 11). This supports the correspondence between movement and audio cues and once again the intimate coupling motion - music.

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 and jerky movements; Sadness with slow and smooth movements; and Fear with small and somewhat irregular movements. However, since the communication of Fear failed, its characterization is questionable.

A slow speed or tempo causes an increase of the movement duration. Similar to the essential role of tempo in music performance, the connection between velocity 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.

Effects of viewing condition (Experiment I)

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 conditions 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.

A somewhat surprising result was the unexpectedly small differences between the different viewing conditions. One possibility 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.

¹ A somewhat ironic fact considering the huge proportion of professional musicians suffering from performance anxiety (see e.g. Fishbein, Middlestadt, Ottati, Straus, & Ellis, 1988).

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 subjects 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 between the 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 players used somewhat different movement characteristics in their performances, an expected result. Not all performers use expressive movements distinctively or in a way that is interpretable for an observer. In fact, many musicians may not appear to move at all. However, Davidson and Correia (2002) points out that 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, Vines, Middleton, McKay, & Hatch, 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 an sound level, only to less extent.

Apart from helping the performer in interpreting the expressive content of the piece there are also other reasons for non-sound generating movements. Some types of movements could actually play a role in facilitating music performance. For instance, it has been shown that body sway can improve reading performance (Stoffregen, Pagulayan, Bardy, & Hettlinger, 2000).

Influence of the instrument

In this study we have studied 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 that the required sound generating movements differ considerably between woodwind instruments, like bassoon and saxophone, and percussion instruments, like marimba, our results suggest that performers on these instruments share a common body language.

To conclude, our results show that specific emotional in-

tentions in music performance can be communicated to observers through movements only. The intentions Happiness, Sadness, and Anger were successfully conveyed while 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.

References

- Abernethy, B., & Russel, 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(1), 21-41.
- Camurri, A., Lagerlöf, I., & Volpe, G. (2003, July). Recognizing emotion from dance movements: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies*, 59(1-2), 213-225.
- Clarke, E. F. (2001). Meaning and specification of motion in music. *Musicae Scientiae*, 5(2), 213-234.
- Clarke, E. F., & Davidson, J. W. (1998). The body in performance. In W. Thomas (Ed.), *Composition-Performance-Reception* (p. 74-92). Aldershot: Ashgate.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second ed.). Academic Press.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (Third ed.). Lawrence Erlbaum Associates.
- Dahl, S. (2000, Sep). The playing of an accent - preliminary observations from temporal and kinematic analysis of percussionists. *Journal of New Music Research*, 29(3), 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(4), 762-776.
- 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* (p. 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*. (p. 237-250). 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 attribution 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), 1-8.
- 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 Acoustic Society of America*, 105(3), 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* (p. 223-248). Oxford University Press.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & 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(6), 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* (p. 309-337). 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(5), 770-814.
- Juslin, P. N., & Sloboda, J. A. (Eds.). (2001). *Music and emotion*. Oxford University Press.
- 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* (Vol. 19, p. 27-44). 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* (p. 756-761). Laurence Erlbaum Associates.
- Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82(2), B51-B61.

- 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(1), 145-158.
- Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology*, 7(4), 733-740.
- 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*. (p. 55-83). Cambridge: Cambridge University Press.
- Sörgjerd, M. (2000). *Auditory and visual recognition of emotional expression in performances of music*. Unpublished thesis, Uppsala University, Dep. of Psychology, 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). Vocal fold physiology: Voice quality control. In O. Fujimura & M. Hirano (Eds.), *A singer's expression of emotions in sung performance*. (p. 217-229). San Diego: Singular Press.
- Todd, N. P. M. (1999). Motion in music: A neurobiological perspective. *Music Perception*, 17(1), 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* (p. 241-253). 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.