

Dept. for Speech, Music and Hearing
**Quarterly Progress and
Status Report**

**Expressiveness of a marimba
player's body movements**

Dahl, S. and Friberg, A.

journal: TMH-QPSR
volume: 46
number: 1
year: 2004
pages: 075-086



**KTH Computer Science
and Communication**

<http://www.speech.kth.se/qpsr>

Expressiveness of a marimba player's body movements

Sofia Dahl and Anders Friberg

Abstract

Musicians often make gestures and move their bodies expressing their musical intention. This visual information provides a separate channel of communication to the listener. In order to explore to what extent emotional intentions can be conveyed through musicians' movements, video recordings were made of a marimba player performing the same piece with four different intentions, Happy, Sad, Angry and Fearful. Twenty subjects were asked to rate the silent video clips with respect to perceived emotional content and movement qualities. The video clips were presented in different viewing conditions, showing different 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 there were cues that the observers used to distinguish between intentions, similar to cues found for audio signals 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. It can even be perceived if the movement is shown just as single points of light fastened to the body and displayed with high contrast to give a kind of discrete-points 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 and Frykholm, 1981), and landing positions of strokes of badminton playing (Abernethy and Russel, 1987). It is also possible to identify the emotional expression in dance and music performances (Walk and Homan, 1984, Dittrich et al., 1996, Sörgjerd, 2000), as well as the emotional expression in every-day arm movements such as drinking and lifting (Pollick et al., 2001, Paterson et al., 2001).

Music has an intimate relationship with movement in several different aspects. The most obvious relation is that all sounds from traditional acoustic instruments are produced by biological motion of humans. 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 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 having an active role in the sound generation, 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, not subordinate to each other. Bearing in mind that music also is a form of communication and that speech and music have many properties in common (Juslin and 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. To separate the influence of each of the aspects suggested by Davidson and Correia on a specific movement may not be possible, but by concentrating on solo performances without an audience, (2) and (3) above may be the dominating aspects and the more extra-musical influences (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 and 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 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 from the artist's gestures rather than from the musical content (Davidson, 1995).

Davidson 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), in her master thesis study, 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 appropriate 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 proposed several movement cues considered important for detecting emotional expression (De Meijer, 1989, 1991, Boone and Cunningham, 1999, see overview in Boone and 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 et al. (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 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 and 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 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 a fast mean tempo, high sound level, staccato articulation, and fast tone attacks, while a Sad performance is characterized by a 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, and as already mentioned above, 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 imagi-

nary sense of motion (e.g. Clarke, 2001, Shove and 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 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 et al. (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) even suggests 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 coming from real objects it is plausible that there is a perceptual effort for all sound sequences to be translated 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 objective was to find out if expressive communication of specific emotions was possible using body movements, and also whether this communication can be described in terms of movement cues (such as fast - slow, jerky - smooth etc.), cues similar to those appearing when listening to music performances. A number of different aspects of musicians' body movements have been identified above. We assume that in this investigation 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 were the following:

1. How successful is the overall communication of each intended emotion?

2. Are there any differences in the communication depending on intended emotion, or what part of the player the observers see?
3. How can perceived emotions be classified in terms of movement cues?

Method

Stimulus Material

A professional percussionist was asked to prepare 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 musical sound 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. The player reported afterwards that she, during these pauses between performances, prepared for the next performance by recalling memories of situations where she had experienced the same emotion. Informal studies of 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), *no-hands* (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 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 mark how they perceived the character of the movements. 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.

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:	fast	-	slow
Fluency:	jerky	-	smooth
Regularity:	irregular	-	regular

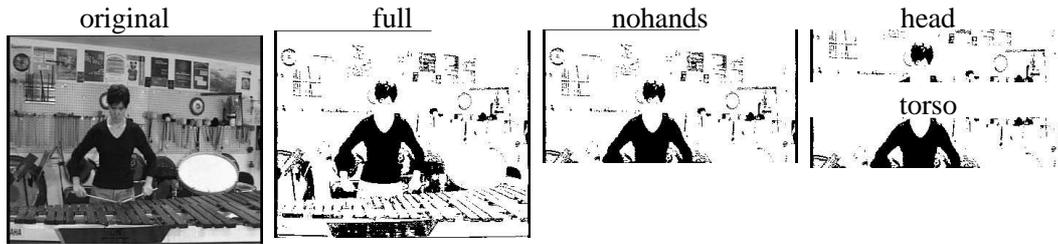


Figure 1: Original (far left) and filtered video images exemplifying the four viewing conditions used in the test: full, nohands, head, and torso.

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 and rated individually. For each subject a batch-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.

Results

Emotion ratings

A measure of how well the intended emotion was communicated to the listener was computed. The *achievement* was defined as the similarity between the *intended* (\mathbf{x}) and the *rated* (\mathbf{y}) emotion, for each video presentation. 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 Happy $\mathbf{x} = [F \ A \ H \ S] = [0 \ 0 \ 1 \ 0]$ and the maximum achievement would be for a rating of $\mathbf{y} = [F \ A \ H \ S] = [0 \ 0 \ 6 \ 0]$. The achievement $A(\mathbf{x}, \mathbf{y})$ for a specific presentation is defined as

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

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. This means that we normalize relatively the best possible case rather than relative the actual data. Given that \mathbf{x}

can only take the values 0 and 1, and \mathbf{y} can be integer values between 0 and 6, $C = 1.125$. A negative achievement value 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 an achievement significantly larger than zero implies that the communication of emotional intent was successful. In practice, the achievement measure is the same as the average of the covariance between the intended and rated emotion for each presented video clip, with a normalization factor included.

There were two main reasons to introduce this new measure of achievement: (a) The influence of the independent factors could be analyzed in one ANOVA, summarizing all four rated emotions, and (b) compared to the more obvious point-biserial correlation the achievement measure reflects also the absolute magnitude of the ratings. For example, for the happy intention ($\mathbf{x} = [0 \ 0 \ 1 \ 0]$) a rating of $\mathbf{y} = [0 \ 0 \ 6 \ 0]$ gives a higher achievement value than $\mathbf{y} = [0 \ 0 \ 2 \ 0]$.

Figure 2 shows the mean achievement for all eight performances presented according to intended emotion, viewing condition and performance. The 95 % confidence intervals are indicated by the vertical error bars. The figure illustrates that the player was able to convey most of the intended emotions to the observers in all viewing conditions. Sadness was most successfully identified, followed by Anger and Happiness. By contrast, Fear was hardly recognized at all.

To reveal the importance of the differences between the intended emotions and viewing conditions, the achievement measures were subjected to a 4 conditions x 4 emotions x 2 performances repeated measures ANOVA. The analysis showed main effects for intended emotion [$F(3, 57) = 33.65, p < 0.0001$], and viewing conditions [$F(3, 57) = 9.54, p < 0.0001$], and significant results for the two-way interactions: viewing condition x emotion [$F(9, 171) = 4.46, p < 0.0001$], and emotion x performance [$F(3, 57) = 2.86, p < 0.05$].

The main effect of emotion was clearly due

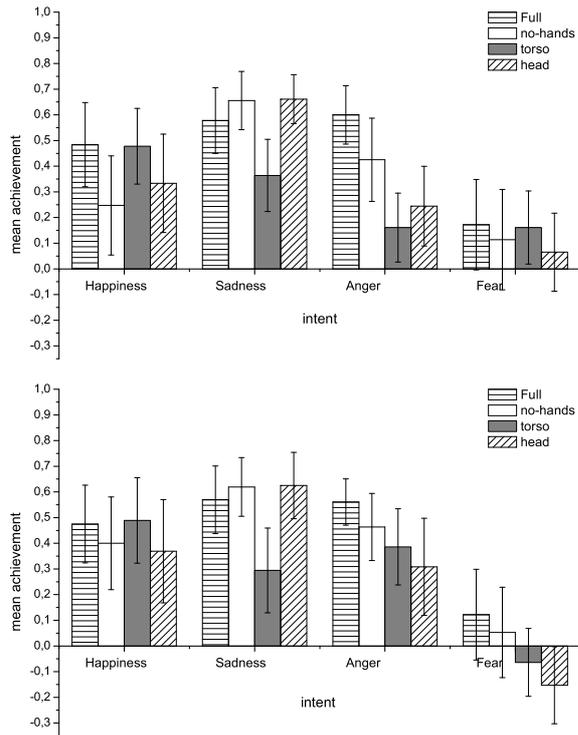


Figure 2: Measure of achievement for the four intended emotions and viewing conditions for the first (top) and second (bottom) performance of each intended emotion. Each bar shows the mean achievement for one emotion and viewing condition; full (horizontally striped), no-hands (white), torso (grey), and head (diagonally striped), averaged across 20 subjects. The error bars indicate 95 % confidence interval. The figures show that the mean achievements were in fact very similar for the players' two performances of each intention. The intention Fear was to a higher extent rated as other emotions in the second repetition, resulting in negative achievement.

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 ($p < 0.02$) achievement than the Sad intention. No other pair-wise comparisons were significant.

Surprisingly, the visible part of the player played less a role than was expected. Although the main effect of viewing condition was significant, the effect was rather small. A Tukey post hoc test showed no significant differences between the four viewing conditions. Initially one would hypothesize that seeing more of the player would provide the observer with more detailed information about the intention. The achievement values would then be ordered from high to low for the three conditions full, no-hands and torso, and similarly for full, no-hands and head. Such

Table 1: Correct identification of the intended emotions in percent for the four viewing conditions, averaged across the two performances for each intention. The values were calculated as the portion of ratings where the intended emotion received the highest rating. The viewing condition receiving the most correct identifications for a specific intention is shown in bold.

	full	no-hands	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
col. mean	67.0	53.3	46.8	51.5	

a “staircase” relation between the viewing conditions was only observed for the intention Anger (see Figure 2).

Some differences due to viewing condition could, however, be observed in the interaction between emotion and viewing condition. The significant effect appears to be due to differences between conditions for the Sad and Angry intentions. For the Sad intention the head seems to be important for perceiving the intended expression. All the conditions where the head is visible (full, no-hands, and head) received high achievement values (from 0.57 to 0.66 in Figure 2), while the achievement values for the torso condition was much lower (0.29 and 0.36). For Anger, the full condition received the highest achievement, while the torso and head conditions were less successful in conveying the intention, particularly in the first performance (compare top and bottom panels in Figure 2).

The significant interaction between emotion and performance could be partly explained by the different ratings for Anger between the two performances. However, also the intention Fear seems to have contributed to this. The second performance of Fear was confused with other emotions to such an extent that it received much more negative achievement values.

The confusion between different emotions can be studied in more detail in Figure 3. Each panel shows the mean ratings for the four emotions averaged across the 20 subjects and the two performances of each intended emotion. The occasional confusion of Anger with Happiness and vice versa indicates that these two expressions might have some features in common. The ratings for the intention Fear, however, are more evenly spread across the four available emotions.

In order to compare our results to other studies the subjects' ratings were transformed into

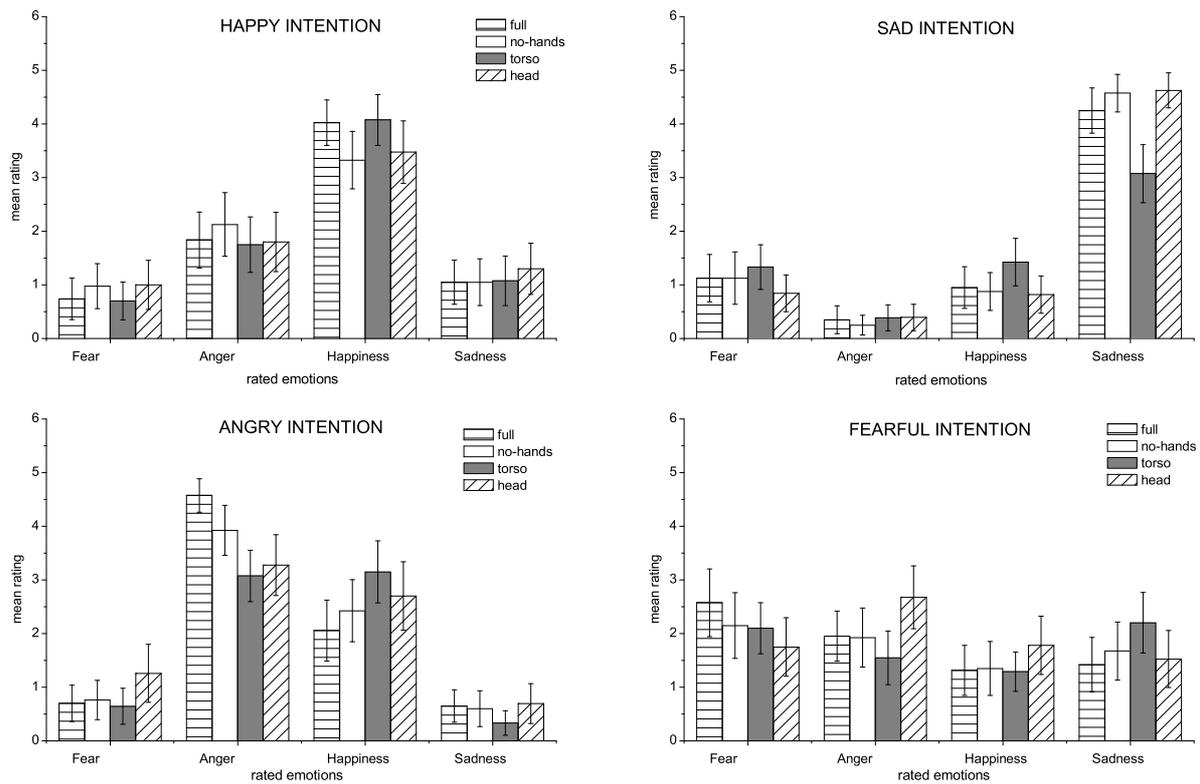


Figure 3: Ratings for the four intended emotions and viewing conditions. Each panel shows the mean ratings for the four emotions averaged across 20 subjects and the two performances of each intended emotion. The patterns of the bars show the four viewing conditions: full (horizontally striped), no-hands (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 Scared (bottom right) was hardly recognized at all.

forced choice responses. 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 1. The pattern of these values corresponds very closely to the mean achievement across the performances shown in Figure 2. Sadness, Anger, and Happiness were identified well above chance level (25%).

Movement cues

Figure 4 shows the mean ratings of the movement cues for each intended emotion. The different movement cues; Amount (none - large), Speed (fast - slow), 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, slow speed, smooth fluency, and regular movements, while low ratings correspond to small amount of movement, fast speed, jerky fluency, and irregular movements.

The intentions Happiness and Anger obtained similar rating patterns. 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. In contrast, the ratings for the Sad performances display small, slow, smooth and regular movements. The ratings for Fear are less clear-cut, but tend to be somewhat small, fast, and jerky.

Table 2 shows the intercorrelations between the movement cues. As expected, they were all somewhat correlated with values ranging from - 0.26 to 0.62. 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 inter-

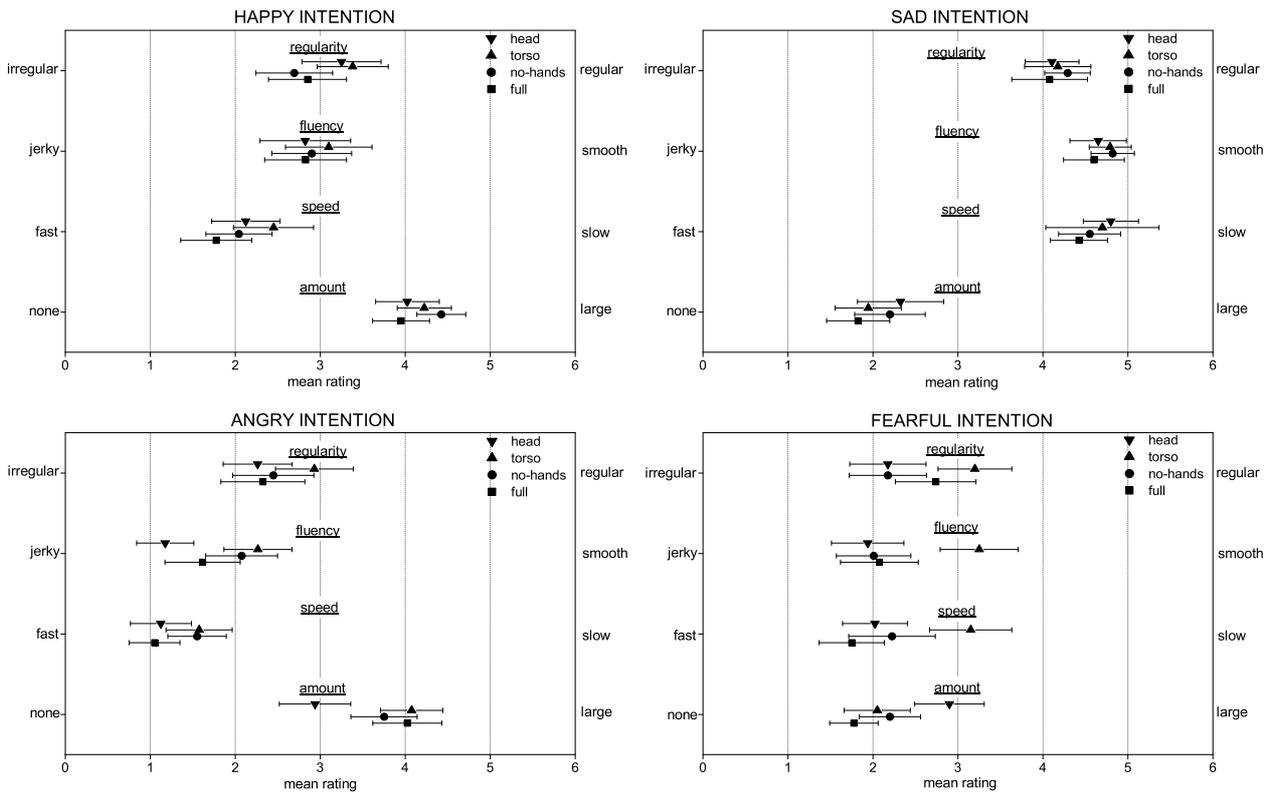


Figure 4: Ratings of movement cues for each intended emotion and viewing condition. Each panel shows the mean markings for the four emotions averaged across 20 subjects and the two performances of each intended emotion. The four viewing conditions are indicated by the symbols: full (square), no-hands (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.

Table 2: Intercorrelations between the movement cues. All correlations were statistically significant ($p < 0.01$, $N = 617$)

	amount	speed	fluency	regularity
amount	-			
speed	-0.26	-		
fluency	-0.19	0.62	-	
regularity	-0.12	0.44	0.58	-

correlations.

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 3 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 low values in terms of explained variance, ranging from 10 to 42%. Applying mul-

tiples 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.

The semipartial correlation sr , was used to estimate the relative importance of each movement cue (shown in italics in Table 3). It expresses the unique contribution from each independent variable, excluding the shared variance (Cohen et al., 2003). According to the table, the cue that was the most important for predicting Happiness was Amount (large, $sr = 0.34$), followed by Speed (fast, $sr = -0.16$). Similarly for Anger the most important cues were Amount (large, $sr = 0.18$), and Fluency (jerky, $sr = -0.21$), and to a lesser degree Speed (fast, $sr = -0.14$), and Regularity (irregular, $sr = -0.13$).

For comparison, four audio cues that are important for identifying the emotions in music performance are shown to the right in Table 3: sound level, tempo, articulation and tempo variability. Note the close correspondence in the comparison

Table 3: Results from the regression analysis for the rated emotions and rated movement cues. 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				audio cues			
	amount	speed	fluency	regularity	sound lev.	tempo	artic.	tempo var.
Happiness	0.35	-0.21	-0.06	-0.01	high	fast	staccato	small
<i>R</i> = 0.44	<i>0.34</i>	<i>-0.16</i>	<i>0.04</i>	<i>-0.01</i>				
Sadness	-0.18	0.43	0.15	0.08	low	slow	legato	final rit.
<i>R</i> = 0.65	<i>-0.18</i>	<i>0.33</i>	<i>0.11</i>	<i>0.06</i>				
Anger	0.19	-0.18	-0.30	-0.16	high	fast	staccato	small
<i>R</i> = 0.61	<i>0.18</i>	<i>-0.14</i>	<i>-0.21</i>	<i>-0.13</i>				
Fear	-0.28	-0.05	-0.17	-0.07	low	fast	staccato	large
<i>R</i> = 0.32	<i>-0.28</i>	<i>0.04</i>	<i>-0.12</i>	<i>-0.05</i>				

between the variation for 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.

Differences in cue ratings for different viewing conditions were, in general, small. For the intentions Happy and Sad and partly for Anger, the cue ratings are closely clustered (see Figure 4). 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 head's movements.

Conclusion and Discussion

The objective for this study was to investigate whether it is possible to convey specific emotions, using body movements only, and if movement cues can be used to describe this communication. Our results show that the intentions Sadness, Happiness, and Anger were successfully conveyed, while Fear was not. The identification of the intended emotion was only slightly influenced by the viewing condition, although in some cases the head was important.

Rated movement cues could be used to characterize the different emotional intentions: Anger was associated with large, fast, irregular, and jerky movements; Happy with large and somewhat fast movements, Sadness with small, slow, and smooth movements, and Fear with somewhat small and jerky movements. However, since the communication of Fear failed, its characterization is questionable.

A strong correspondence was found between how the selected movement cues and their audio counterparts varied (see Table 3). This supports the correspondence between the movement and

audio cues and once again the intimate coupling motion - music.

The most successfully conveyed intention was Sadness. For the Sad intention it was also evident that the movements of head provided important cues for correctly identifying the intention. The viewing conditions where the head was not visible (torso) obtained lower ratings than the other conditions. The low Sadness ratings for the torso condition occurred for both performances of the Sad intention (see Figure 2). A possible explanation could be that for this intention there is a specific cue from the player visible in the head only. Our visual inspections of the head stimuli suggest that for the Sad performances the head was mostly turned downwards. Also, there were less and slower movements in the vertical direction compared to the other intentions.

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 was a clear change in the classification and intensity ratings for all three intentions. 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 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 and Homan, 1984, Dittrich et al., 1996, Boone and 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. That Sadness was easily communicated in the present study is not surprising considering that children as young as four years old are able to identify Sad performances (Boone and Cunningham, 1998), and also produce the relevant cues to have adult observers identify Sadness at above chance level (Boone and 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 was also found by Dittrich et al. (1996) when presenting subjects with point-light dance performances. When the dances were showed in normal lighting the intention Joy was more likely to be mistaken for Surprise.

A 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 no-hands 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 found that the hands provided no informa-

tion 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.

We have been studying performances by a single player only, but still it could be of some interest to discuss the possible meaning of the body language of percussion players. Strictly speaking the player 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.

Acknowledgement

The authors would like to thank Alison Eddington for the marimba performances, all persons participating as subjects in the viewing test, 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)

References

- Abernethy, B. and 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. and 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. and Cunningham, J. G. (1999). The attribution of emotion to expressive body movements: A structural cue analysis. Manuscript submitted for publication.
- Boone, R. T. and 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., and Volpe, G. (2003). 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. and Davidson, J. W. (1998). The body in performance. In Thomas, W., editor, *Composition–Performance–Reception*, pages 74–92. Aldershot: Ashgate.
- Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, third edition.
- Dahl, S. (2000). 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 Steinberg, R., editor, *Music and the mind machine: Psychophysiology and psychopathology of the sense of music*, pages 105–114. Heidelberg: Springer.
- Davidson, J. W. and Correia, J. S. (2002). Body movement. In Parncutt, R. and McPherson, G. E., editors, *The science and psychology of music performance. Creative strategies for teaching and learning.*, pages 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., and Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, 6:727–738.
- Friberg, A. and 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. and 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.
- 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 Juslin, P. and Sloboda, J. A., editors, *Music and Emotion*, pages 309–337. Oxford University Press.
- Juslin, P. N., Friberg, A., and 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. and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5):770–814.
- McNeill, D., Quek, F., McCullough, K.-E., Duncan, S., Bryll, R., Ma, X.-F., and Ansari, R. (2002). Dynamic imagery in speech and gesture. In Granström, B., House, D., and Karlsson, I., editors, *Multimodality in Language and Speech Systems*, volume 19 of *Text, Speech and Language Technology*, pages 27–44. Kluwer Academic Publishers.
- Paterson, H. M., Pollick, F. E., and Sanford, A. J. (2001). The role of velocity in affect discrimination. In Moore, J. D. and Stenning, K., editors, *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society, Edinburgh*, pages 756–761. Laurence Erlbaum Associates.
- Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82(2):B51–B61.
- Runeson, S. and Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology*, 7(4):733–740.
- Shove, P. and Repp, B. (1995). Musical motion and performance: theoretical and empirical perspectives. In Rink, J., editor, *The practice of Performance. Studies in Musical Interpretation.*, pages 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.
- Todd, N. P. M. (1999). Motion in music: A neurobiological perspective. *Music Perception*, 17(1):115–126.
- Walk, R. D. and 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 Wachsmuth, I. and Sowa, T., editors, *Gesture and Sign Language in Human-Computer Interaction*, pages 241–253. April, Springer Verlag.