

## WHAT CAN THE BODY MOVEMENTS REVEAL ABOUT A MUSICIAN'S EMOTIONAL INTENTION?

*Sofia Dahl and Anders Friberg*

Department of Speech Music and Hearing  
Kungl Tekniska Högskolan  
sofia@speech.kth.se, andersf@speech.kth.se

### ABSTRACT

Music has an intimate relationship with motion in several aspects. Obviously, movements are required to play an instrument but musicians move also their bodies in a way not directly related to note production. In order to explore to what extent emotional intentions can be conveyed through musicians' movements only, video recordings of a marimba player performing the same piece with the intentions Happy, Sad, Angry and Fearful, were recorded. 20 observers watched the video clips, without sound, and rated both the perceived emotional content as well as movement cues. The videos were presented in four viewing conditions, showing different parts of the player. The observers' ratings for the intended emotions 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 the viewing condition, although in some cases the head was important. The movement ratings indicate that there are cues that the observer use to distinguish between intentions, similar to the cues found for audio signals in music performance. Anger was characterized by large, fast, uneven, and jerky movements; Happy by large and somewhat fast movements, Sadness by small, slow, even and smooth movements.

### 1. INTRODUCTION

Musical performances are often enjoyed visually as well as aurally. It is not unusual to see the audience' necks stretched in attempt to follow the musicians' movements. That it would only be the actual sound producing movements that interest us seems unlikely, for these movements are often too small or too fast to be seen properly. However, musicians move also in ways that are not directly related to the production of notes. These movements have been shown to be able to convey information about the expressive intent of performances. For instance, in studies by Davidson [1][3], subjects were about equally successful in rating music performances according to their expressive intent (deadpan, projected or exaggerated) regardless if they were allowed to only listen, only watch, or both watch and listen. The musically naive were even better in recognizing the intent in the watch-only mode, compared to the other modes [3].

The ability of observers to obtain information regarding emotional intent (affect) from movements only has been well documented, not only for music performances but also for other settings, such as dancing [8], drinking, or knocking [4]. Work has also been dedicated to what kinds of movement characteristics that provide the pieces of information that observers use in order to distinguish between performances. Some suggestions of

such *movement cues* have been made by DeMeijer and Boone and Cunningham [5] [6] [7]. For instance, actors' movements were associated with Joy when their movements were fast, upward directed, with arms raised, whereas the optimal movements for Sadness were slow, light downward directed, with arms closed around the body [5] [6].

That the direction of movement and the position of the arms seem to be of such importance is interesting in perspective of Davidson's work. Musicians' arm and hand movements are primarily involved in the sound production, and expressive movements used by observers to discriminate between performances must therefore either appear in other parts of the body, or *coincide* with the actual playing movements. Davidson [2] found that observers were not able to identify the expressive intention from the hand movements only, while the head movements seemed to be of greater importance.

In analysis of music performances audio cues, such as tempo, sound level etc., have been found to characterize emotional coloring [9][10]). 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, then, to assume that the body movements in the performances contain cues corresponding to those appearing in the audio signal.

The questions for this study 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? and (3) How can perceived emotions be classified in terms of movement cues?

### 2. EXPERIMENT

A professional percussionist was video recorded when performing a short piece of music with the intentions Sadness, Anger, Happiness and Fear, on the marimba. 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 for the different interpretations.

From the video recordings, stimuli clips were generated showing different parts of the player in four *viewing conditions*: *full* (showing the full image), *nohands* (the player's hands not visible), *torso* (the player's hands and head not visible), and *head* (only the player's head visible). A video editing software was used to cut out the stimuli clips for the four viewing conditions using a cropping

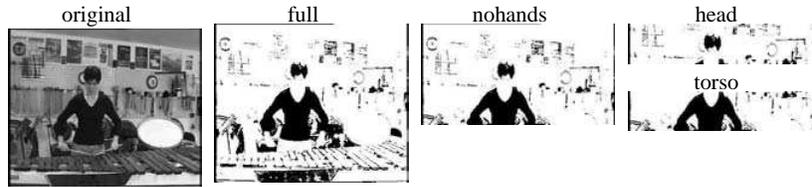


Figure 1: Original (far left) and filtered video images exemplifying the four viewing conditions used in the test: full, nohands, head, and torso.

filter. A threshold filter was also used so that facial expressions would not be visible (see Figure 1). Based on the original eight video recordings a total of 32 (4 emotions x 2 performances x 4 conditions) video clips were generated. The duration of the video clips varied between 30 and 50 s.

Twenty subjects watched the video clips individually and rated the emotional content on a scale from 0 (nothing) to 6 (very much) for the emotions Fear, Anger, Happiness, and Sadness. The subjects were also asked to mark how they perceived the movements. The ratings were done on bipolar scales (from 0 to 6) for the cues:

Amount: none - large  
 Speed: fast - slow  
 Fluency: jerky - smooth  
 Distribution: uneven - even

### 3. RESULTS

#### 3.1. Measure of achievement

From the 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

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

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. In this case, 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.

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, Happiness and Anger were all well communicated, while Fear received low, sometimes negative, achievement.

In order to facilitate comparisons with other results the proportion of correct identifications were calculated by converting the ratings to "forced choice" answers. The conversion was made strictly, meaning that only the answers were the intended emotion received the highest ratings were considered as "correctly" identified. The proportion of correct identifications for each intended emotion are indicated by the small black squares above each bar in Figure 2. The proportion of correct responses follow the same pattern as achievement, with the highest values for the intention Sadness (95% correct), followed by Anger, Happiness, and Fear. Despite the fact that the responses where the intended emotion was rated equal to another emotion were treated as "incorrect", the correct identifications are well above chance level (25%) in most cases.

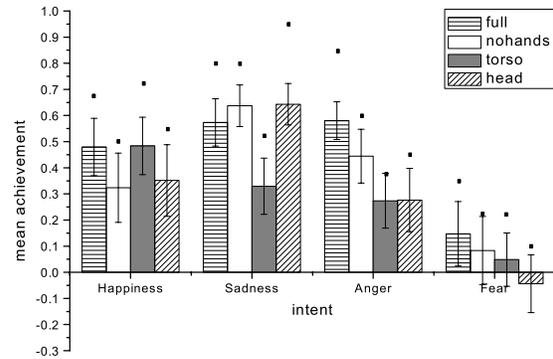


Figure 2: Mean achievement for the four intended emotions and viewing conditions averaged across the first and second performance of each intended emotion. Each bar shows the mean achievement for one emotion and viewing condition, full (horizontally striped), nohands (white), torso (grey), and head (diagonally striped), averaged across 20 subjects and two performances. The error bars indicate 95% confidence interval. Performances with the intentions Happiness, Sadness, and Anger received ratings in correspondence with the intention, while the Fearful performances were hardly recognized at all. Above each bar a small black square indicate the relative proportion of correctly identifications, as calculated from the highest rated emotion for each stimulus response.

### 3.2. Influence of viewing conditions and emotions

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 [ $F(3, 57) = 2.86, p < 0.05$ ].

Although the main effect of viewing condition was significant, the effect was surprisingly small, see Figure 2. 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, nohands and head, and similarly for full, nohands and torso. Such a "staircase" relation between the viewing conditions was only observed for the intention Anger.

The significant interaction between emotion and viewing condition seems to be due to differences in the Sad and Angry intention. For the Sad intention the head seems to be of highest importance in perceiving the intended expression. All the conditions where the head is visible (full, nohands, and head) received high ratings for Sadness with mean achievements from 0.57 to 0.64, while torso rated a much lower mean achievement of 0.32. For Anger, the full condition received the highest Anger ratings, while the conditions torso and head seem less successful in conveying the intention, particularly in the first performance.

Overall, the mean achievements proved to be very similar for the player's two performances of each intention. The intention Fear was rated as other emotions to a higher extent in the second performance, resulting in negative achievement.

### 3.3. Movement cues

Figure 3 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 Distribution (uneven - even), received different ratings depending on whether the intended expression was Happy, Sad, Angry, or Fearful. Note that high ratings correspond to large amounts of movement, slow speed, smooth fluency, and even distribution, while low ratings correspond to small amounts of movement, fast speed, jerky fluency, and uneven distribution.

The intentions Happiness and Anger obtained similar rating patterns. Both Anger and Happiness seem to display large movements, but the Angry performances are somewhat faster and jerkier compared to the Happy performances. In contrast the ratings for the Sad performances display small, slow, smooth and even movements. The ratings for Fear are less clear-cut, but tend to be somewhat small, fast, and jerky. A similar pattern was found when investigating how the subjects related the emotions to the movement cues. The correlation between the rated emotions and the ratings of movement cues is shown in Table 1. According to the table, Anger is associated with large, fast, uneven, and jerky movements; Happy with large and somewhat fast movements, Sadness with small, slow, even and smooth movements, and Fear with somewhat small, jerky and uneven movements. However, since the communication of Fear failed, its characterization is questionable.

Differences in cue ratings for different viewing conditions were, in general, small. For the intentions Happy and Sad and partly for

	amount	speed	fluency	distrib.
Happiness	0.40	-0.27	-0.15	-0.12
Sadness	0.32	0.60	0.50	0.38
Anger	0.31	-0.48	-0.54	-0.44
Fear	-0.24	-0.01	-0.13	-0.11

Table 1: Correlations between rated emotions and rated movement cues. All correlations, except between Fear and speed, were statistically significant ( $p < 0.01, N = 603$ ).

Anger, the cue ratings are closely clustered. 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.

## 4. DISCUSSION

The results show that the four intended emotions were communicated successfully, with the exception of Fear. The most successfully conveyed emotion seems to be Sadness.

While there generally were surprisingly small differences between viewing conditions, the head seemed to be very important for correctly identifying the Sad intention. The only viewing condition where the head was not visible, torso, received much lower Sadness ratings than the other conditions for both performances with the Sad intention. Our visual inspections of the stimuli clips revealed no extraordinary features in the movement of the head, but for the Sad performances there do seem to be less and slower movements in the vertical direction compared to the other intentions.

Our results for the ratings of movement cues resemble the cues used by young children in the study by Boone and Cunningham [8]. They reported that the children used more force and rotation and a higher tempo when portraying Happiness and Anger than they did for Sadness and Fear. Their cues force and rotation correspond well to our cues for amount of movement and speed. The children also used fewer shifts in movement patterns for Sadness than for the other emotions, something that bears similarities to our cues for fluency and distribution.

There is also a strong resemblance between these movement cues and the audio cues used in expressive music performances. The most evident connection seem to be between movement speed and musical tempo, but also the similarities between amount of movement and sound level, or fluency and articulation, seem clear.

## 5. CONCLUSIONS

Our results show that the intentions Sadness, Happiness, and Anger were conveyed through musician's movements only, 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.

The movement cues used in the communication have similarities to the cues found for audio signals in music performance. Anger was characterized by large, fast, uneven, and jerky movements; Happy by large and somewhat fast movements, Sadness by small, slow, even and smooth movements. Further research could reveal whether the movement cues reported here would apply also for other performers and instruments.

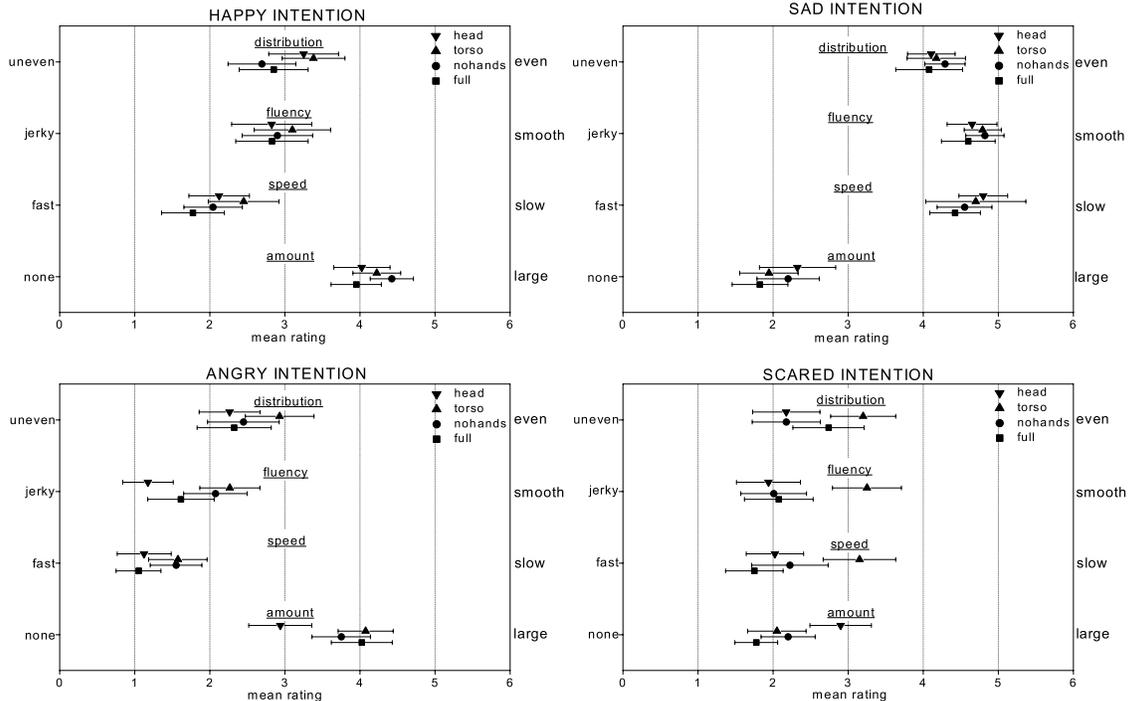


Figure 3: Ratings of movement cues for each intended emotion and viewing condition. Each panel shows the mean markings for the four emotions averaged across 20 subjects and the two performances of each intended emotion. The four viewing conditions are indicated by the symbols: full (square), nohands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

## 6. ACKNOWLEDGEMENTS

The authors would like to thank Alison Eddington for the marimba performances and all persons participating as subjects in the viewing test. This work was supported by the European Union (MEGA - Multisensory Expressive Gesture Applications, IST-1999-20410; <http://www.megaproject.org/>)

## 7. REFERENCES

- [1] Davidson, J. W., "Visual perception and performance manner in the movements of solo musicians", *Psychology of Music*, Vol. 21, 1993, 103–113.
- [2] Davidson, J. W., "What type of information is conveyed in the body movements of solo musician performers?", *Journal of Human Movement Studies*, Vol. 6, 1994, 279–301.
- [3] Davidson, J. W., "What does the visual information contained in music performances offer the observer? Some preliminary thoughts", In Steinberg, R. (Ed.) *Music and the mind machine: Psychophysiology and psychopathology of the sense of music*, Heidelberg: Springer, pp. 105–114, 1995.
- [4] Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J., "Perceiving affect from arm movement" *Cognition*, 82(2), 2001, pp. B51-B61.
- [5] De Meijer, M., "The contribution of general features of body movement to the attribution of emotions" *Journal of Nonverbal Behavior*, Vol. 13, 1989, 247–268.
- [6] De Meijer, M., "The attribution of aggression and grief to body movements: The effects of sex-stereotypes" *European Journal of Social Psychology*, Vol. 21, 1991, 249–259.
- [7] Boone, R. T., and Cunningham, J. G., "Children's decoding of emotion in expressive body movement: The development of cue attunement", *Developmental Psychology*, Vol. 34, 1998, 1007–1016.
- [8] Boone, R. T., and Cunningham, J. G., "Children's expression of emotional meaning in music through expressive body movement", *Journal of Nonverbal Behavior*, 25(1), 2001, 21–42.
- [9] Gabrielsson, A., and Juslin, P. N., "Emotional expression in music performance: Between the performer's intention and the listener's experience", *Psychology of Music*, Vol. 24, 1996, 68–91.
- [10] Juslin, P. N., "Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception", *Journal of Experimental Psychology: Human Perception and Performance*, 26(6), 2000, 1797-1813.