Expressiveness of musician's body movements in performances on marimba.

Dahl, Sofia; Friberg, Anders

Published in:
Lecture Notes in Computer Science

Publication date:
2004

Document Version
Accepted author manuscript, peer reviewed version

Expressiveness of musician's body movements in performances on marimba

Sofia Dahl and Anders Friberg

Dept. of Speech, Music and Hearing
KTH, Royal Institute of Technology
SE-100 44 Stockholm, Sweden
sofia@speech.kth.se, andersf@speech.kth.se

Abstract. 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 the intentions Happy, Sad, Angry and Fearful. 20 subjects were presented video clips, without sound, and asked to rate both the perceived emotional content as well as the movement qualities. The video clips were presented in different conditions, showing the player to different extent. The observers' ratings for the intended emotions confirmed that the intentions Happiness, Sadness and Anger were well communicated, while Fear was not. Identification of the intended emotion was only slightly influenced by the 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.

1 Introduction

Music has an intimate relationship with movement in several different aspects. First of all, all sounds on traditional acoustic instruments are produced by human motion. Some characteristics of this motion will inevitably be reflected in the resulting tones.

Musicians move also their bodies in a way that is not directly related to the production of notes. Wanderley [1] refers to these other performer movements as *ancillary, accompanist* or *non-obvious* movements. We prefer to think of it as a kind of *body language* since, as we will see below, it serves several important functions in music performance. It seems reasonable to assume that some of the expressivity in the music is reflected in these movements. However, the body movements may also be used for more explicit communication. Davidson and Correia [2] suggests 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 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 now 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. [3][4][5]). It was found that subjects were about equally successful in identifying the expressive intent in piano or violin performances, regardless of whether they were allowed to only listen, only watch or both watch and listen [3]. Musically naive subjects actually performed better when only watching, compared to the other conditions [6]. In a similar study, focusing on emotional expression, Sörgjerd [7] found that subjects were more successful in identifying the emotions Happiness, Sadness, Anger and Fear, than the emotions Solemnity and Tenderness.

An interesting comparison can be done between the studies mentioned and how musical expressiveness is encoded and decoded using only the sound. In analysis of music performances Gabrielsson and Juslin [8][9] found that there are a number of cues (such as tempo, sound level etc) that listeners utilize when discriminating between different emotional expression 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.

In this study the objective was to explore both the communication of more specific expressive intentions, and whether this communication can be described in terms of motion cues (such as fast - slow, abrupt - 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 (1) movements for the direct sound production on the instrument, and (2) natural expressive movements not primarily intended to convey visual information to the audience or to fellow musicians.

The questions for this study were the following:

1. How good is the overall communication of each intended emotion?
2. Are there any differences 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?

2 Recording

A professional percussionist was asked to prepare a piece for marimba with four different expressive intentions: Anger, Happiness, Sadness and Fear. 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.

The recording was carried out using a digital video camera (SONY DCR-VX1000E). 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 break between each performance.

3 Experiment

Stimuli. The original video recordings were presented to the observers in different *viewing conditions*, showing the player in the image 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).

Figure 1 shows an example of a frame with the four viewing conditions. The original video files were edited using a freeware video editing software (Virtual Dub). To remove facial expression a threshold filter of 19 % was used, transforming the color image to a strict black and white image (without gray scales). The four conditions were then cut out from the original full scale image, using a cropping filter. Based on the original eight video recordings a total of 32 (4 emotions x 2 repetitions x 4 conditions) video clips were generated.

Subjects and Method. Twenty (10 male and 10 female) subjects participated in the experiment, mostly students and researchers at the department. The subjects did not receive any compensation for their participation.

Subjects were asked to rate the emotional content for each video clip 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 movements. The ratings of movement cues were carried out using bipolar scales. The scales for movement ratings were amount of movement (no movement - large movements), speed (fast - slow), fluency (jerky - smooth) and the distribution of movements (uneven - even).

The 32 video clips were shown on a PC and rated individually. Each clip could be viewed as many times as the subject liked, but subjects could not go back to rate the previous one.

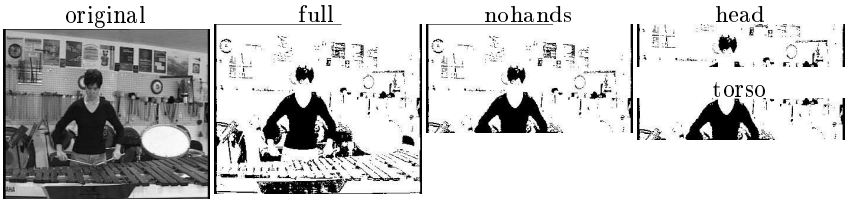

Fig. 1. *Original (far left) and filtered video images exemplifying the four viewing conditions used in the test: full, nohands, head, and torso.*

4 Results and Discussion

Emotion Ratings. In Fig. 2 the mean ratings for each of the intended emotions and the viewing conditions can be seen. Each panel shows the mean ratings for the four emotions averaged across 20 subjects and the two performances of each intended emotion. The Sad intention was the most successfully identified, followed by Angry and Happy. The occasional confusion of Anger with Happiness and vice versa indicate that these two expressions might have some features in common. The ratings for intention Fear, however, seems to be evenly spread across the four available emotions.

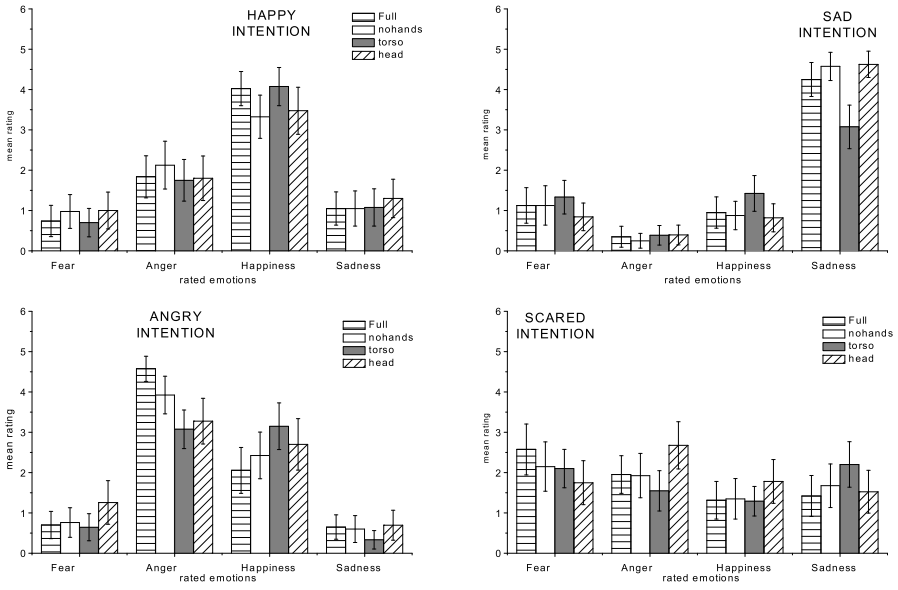

Fig. 2. Ratings for the four intended emotions and viewing conditions. Each panel shows the mean ratings for the four emotions averaged across 20 subjects and the two performances of each intended emotion. The pattern of the bars show the four viewing conditions: full (horizontally striped), nohands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

The influence of the different viewing conditions on the ratings is surprisingly small, but some interaction effects with the intended emotion can be observed.

For the Sad intention all the conditions where the head is visible (full, nohands, and head) receive high ratings for Sadness (means from 4.3 to 4.6 in Fig. 2), while torso rates much lower (mean 3.1). For Anger, the full condition receives the highest Anger ratings, while the conditions torso and head seem less successful in conveying the intention.

By transforming the ratings into forced choice responses, the subjects' identification of the intended emotions were calculated. In doing this, only the ratings where the intended emotion received the highest rating were considered as "correct". Responses where several emotions were rated equally high were considered as incorrect. The percentage of correct responses are shown in Table 1. The pattern of these values corresponds well to the mean ratings across the performances shown in Fig. 2.

The percentage correct responses also relates well to comparisons with other studies. Subjects in this study performed equally well, or better in terms of percentage correct identifications (c.f. [10][11][12]).

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	nohands	torso	head
Happiness	68	50	**73**	56
Sadness	80	80	53	**95**
Anger	**85**	60	38	45
Fear	**35**	23	23	10

Movement Cues. Figure 3 shows the mean ratings for movement cues for each intended emotion. The different movement cues; amount of movement (none - large), speed (fast - slow), fluency (jerky - smooth) and movement distribution (uneven - even), received different ratings depending on whether the intended expression was Happy, Sad, Angry, or Scared. 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 2. According to the table, Anger is associated with large, fast, uneven, and

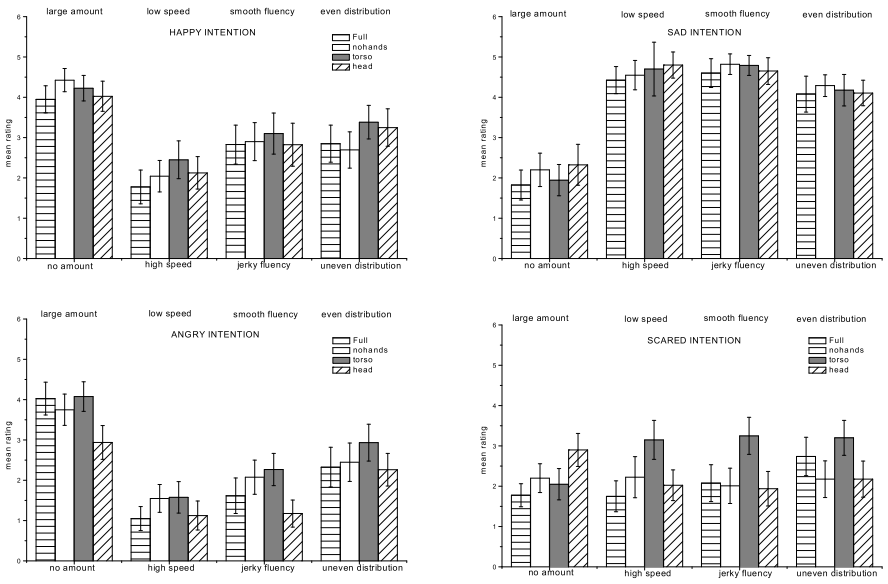

Fig. 3. Ratings for movement cues for each intended emotion and viewing condition. Each panel shows the mean markings for the four emotions averaged across 20 subjects and the two performances of each intended emotion. The pattern of the bars show the four viewing conditions: full (horizontally striped), nohands (white), torso (grey), and head (diagonally striped). The error bars indicate 95 % confidence intervals.

Table 2. Correlations between rated emotions and rated movement cues. All correlations, except between Fear and speed, were statistically significant ($p < 0.01, N = 603$).

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

Table 3. Intercorrelations between the movement cues. All correlations were statistically significant ($p < 0.01, N = 617$).

	amount	speed	fluency	distrib.
amount	-			
speed	-0.26	-		
fluency	-0.19	0.62	-	
distrib.	-0.12	0.44	0.58	-

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 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. Also Davidson [4] found that the head was important for the observers to discriminate between expressive performances, while the hands were not.

In order to check the independence of the different cues the intercorrelations were calculated, see Table 3. As expected, they are all somewhat correlated with values ranging from -0.26 to 0.62. The amount of movement seems to be relatively independent since the intercorrelations with the other cues are rather small. Speed, Fluency and Distribution all show relatively medium intercorrelations.

5 Conclusions

The communication of the four intended emotions were successful, with the exception for Fear. The most successfully conveyed intention seems to be Sadness. For this intention it was also evident that the head provided important cues for correctly identifying the intention. The viewing conditions where the head was not visible (torso) got much lower ratings than did the other conditions. Possible explanations could be that there is a specific cue from the player occurring for this intention, visible in the head only, that observers have learned to recognize.

The movement ratings indicate that there was less amount of movement for Sadness and Fear than for Anger and Happiness; lower speed and more smooth and even movements for Sadness than for Happiness and Fear, which in turn had lower speed with more smooth and even movements than Anger. The movement ratings indicate that there are cues in expressive movements that bear strong resemblance to the audio cues used in expressive music performances. In music performance Sadness is typically characterized by slow tempo, low sound level and legato articulation, while Anger manifests itself through high sound level, fast tempo and abrupt sound onsets. The connection between movement speed and musical tempo seem rather obvious, but also the similarities between amount of movement and sound level, or fluency and articulation, seem clear. Further research could reveal whether these cues are the same for other performances and instruments.

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/)

References

[1] Wanderley, M. M., "Quantitative Analysis of Non-Obvious Performer Gestures", In Wachsmuth, I. and Sowa, T. (Eds.) Gesture and sign language in Human-Computer Interaction, Springer Verlag (2002) 241–253.

[2] Davidson, J. W., Correia, J. S., "Body movement" In Parncutt, R., McPherson, G. E. (Eds.) The science and psychology of music performance. Creative strategies for teaching and learning. Oxford University Press (2002) 237–250.

[3] Davidson, J. W.,"Visual perception and performance manner in the movements of solo musicians", Psychology of Music, **21** (1993) 103–113.

[4] Davidson, J. W.,"What type of information is conveyed in the body movements of solo musician performers?", Journal of Human Movement Studies, **6** (1994) 279–301.

[5] Clarke, E. F., Davidson, J. W., "The body in performance", In Thomas, W. (Ed.) Composition-performance-reception. Aldershot, Ashgate, (1998), 74–92.

[6] 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 (1995) 105–114.

[7] Sörgjerd, M., "Auditory and visual recognition of emotional expression in performances of music", Unpublished thesis, Uppsala University, Dep of Psychology, Uppsala Sweden (2000).

[8] Gabrielsson, A., and Juslin, P. N., "Emotional expression in music performance: Between the performer's intention and the listener's experience", Psychology of Music, **24** (1996) 68–91.

[9] Juslin, P. N., "Communicating emotion in music performance: A review and theoretical framework", In Music and Emotion. Oxford Unviersity Press (2001) 309–337. Journal of Experimental Psychology: Human Perception and Performance, **26(6)** (2000) 1797-1813.

[10] Walk, R. D, Homan, C. P., "Emotion and dance in dynamic light displays represented in dance", Bulletin of Psychonomic Society **22** (1984) 437–440.

[11] Dittrich, W. H., Troscianko, T., Lea, S. E., and Morgan, D., "Perception of emotion from dynamic point-light displays represented in dance", Perception **6** (1996) 727–738.

[12] Camurri, A., Lagerlöf, I., Volpe, G., "Recognizing emotion from dance movements: Comparisons of spectator recognition and automated techniques", International Journal of Human-Computer Studies **59** (2003), 213–225.