Production and perception of gestures in music performance.

Dahl, Sofia

Publication date:
2006

Document Version
Accepted author manuscript, peer reviewed version

PRODUCTION AND PERCEPTION OF GESTURES IN MUSIC PERFORMANCE.

Sofia Dahl
Cognitive and Systematic Musicology Lab, School of Music, Ohio State University
&
Speech, Music and Hearing KTH, School of Computer Science and Communication, Stockholm

ABSTRACT

Performers strive to acquire playing techniques that will allow them to perform under varying playing conditions. However, performing specialized, complex motor tasks under a strict time constraint is not enough. For a performance to be truly successful it should also convey something to the listener, to be expressive. To convey an expressive intention a performer does not only relay on sound, also the visual channel is utilized. The movements of a musicians are both intended to control their instrument and for communicative purposes. This paper concerns both these aspects of musical gestures. I will briefly report work on playing gestures in drumming, as well as observers' visual perception of specific emotional intentions in music performance.

Recordings were made of four drummers' movement strategies when playing accented single strokes for different playing conditions. Analysis of the players' preparatory movements, striking velocity, and timing patterns showed that the players maintained their playing strategies for different tempi, striking surfaces and dynamic levels. Between players, however, there were considerable differences in movement patterns and emphasis on the accented stroke. The dynamic level influenced both preparatory movements, and striking velocity. All players initiated the accented strokes from a greater height, and delivered the accent with increased striking velocity compared to the unaccented strokes. Despite initiated from a greater height, the accented stroke was delivered on time.

One marimba player, a saxophonist and a bassoon player were filmed while performing melodies with four emotional intentions; Anger, Happiness, Sadness, and Fear. Video clips (without sound) were shown to observers who rated each clip with respect to emotional expression and movement characteristics. Results showed that the communication of the intentions Anger, Happiness and Sadness was communicated to the observers while Fear was not. The results of the rated movement cues showed similarities to the audio cues reported for music performance. Anger was characterized by jerky movements; Happy by large and somewhat fast and jerky movements; and Sadness with slow and smooth movements. The similarities between audio and visual cues for music performance bring up interesting questions as to the origin and purpose of these gestures.

INTRODUCTION

Movement gestures in music performance can be divided into two broad categories: 1) *control gestures*, primarily intended for the production of notes, indirectly conveying intention and expression through the resulting sound events; and 2) *communicative gestures*, directly expressing intentions of the performer to observers and co-performers. Examples of control gestures are bowing gestures or preparatory movements moving between keys on a piano. Examples of communicative gestures would be head nods or raising arms to sign transition points in the music. There are also movement gestures that fit equally well into both these categories. For instance, some gestures primarily intended for note production can also convey important cues for observers and co-performers. Similarly, there may be gestures that, although they are not directly producing sound, play an important role in facilitating the sound producing gestures and the performance. One related example of such performance-facilitating movements would be body sway, which has been shown to facilitate reading performance (see Stoffregen et al, 2000)

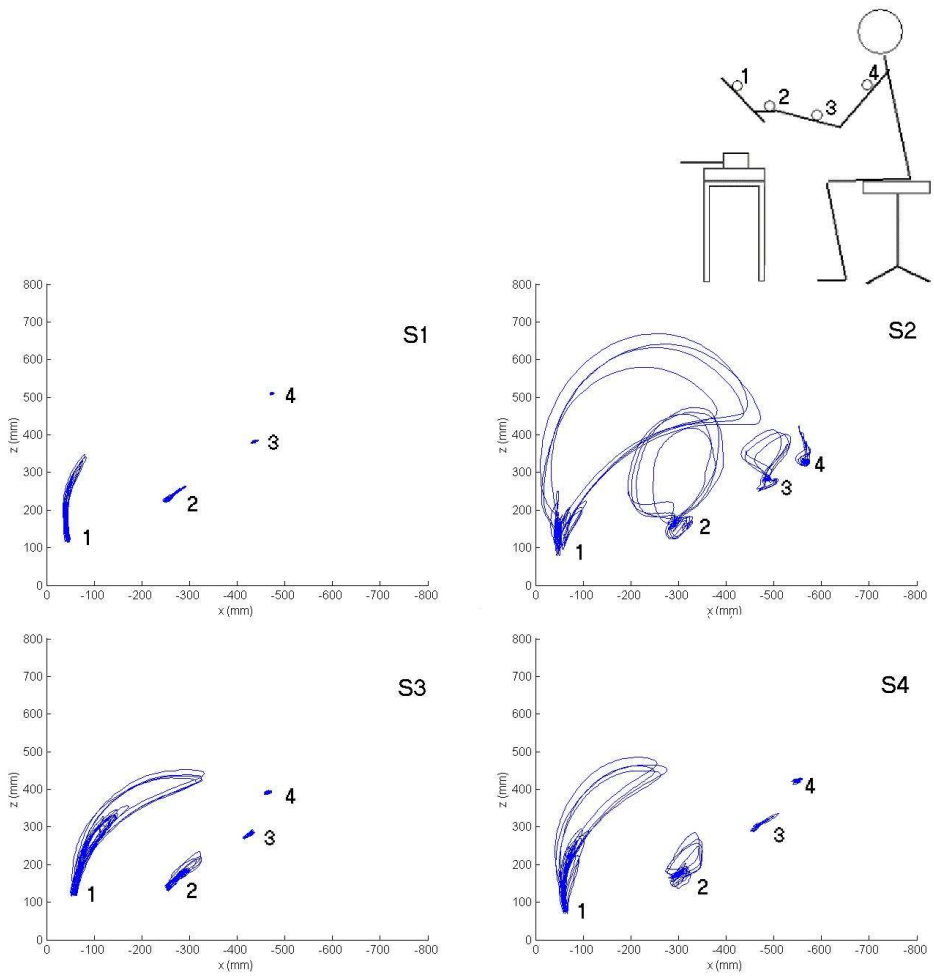
The overlaps between control and communicative gestures make the modeling of movement gestures in music performance somewhat difficult. The separation of general features, appearing across performers, from performer-specific gestures and non-intentional, random variations is a challenge similar to that of distinguishing between expressive timing and random variations in music performances. In order to isolate important features it may be necessary to restrict studies to simplified musical settings.

GESTURE PRODUCTION IN DRUMMING

As a study of the low-level, most automated, aspects of music performance, four percussion players' movement strategies for performing an accented stroke were studied (Dahl, 2004). An accent is an example of a small, but musically relevant, element of music performance which receives a lot of attention in percussion training.

Three professional percussionists and one amateur played on a force plate with markers on the drumstick, hand, and lower and upper arm. The movement trajectories were recorded using a motion detection system (Selspot), tracking the displacement of the markers at 400 Hz. The rhythmic pattern – an ostinato with interleaved accents every fourth stroke – was performed at three dynamic levels (*pp*, *mf*, and *ff*), at three tempi (116, 160, and 200 BPM), and on three different striking surfaces (soft, normal, and hard).

Figure 1: Movement trajectories captured from four markers: on the drumstick, and on the subjects' hand, lower, and upper arm. Side view from the players' left side. Each panel includes approximately four measures at *mf*, 200 beats per minute. The preparatory movements for the accented stroke can be seen as a larger loop compared to that of the unaccented strokes. The players' drumstick, hand, lower, and upper arm are involved to different extent in the movements (Dahl, 2004).



The analysis of the movement trajectories showed that the four players used individual movement strategies to prepare for the accented stroke. Figure 1 displays the movement trajectories of the four markers on the subjects' drumstick, hand, lower and upper arm. The larger loops mark the preparation for the accented strokes. The four players all initiated the accented stroke from a greater height compared to the unaccented strokes. However, they clearly displayed different movement strategies in their preparatory movements both with respect to the phase between limb segments and to what extent the arm was used in the movement (see Figure1). Subject S2 displayed the largest difference in movement trajectories between the two types of strokes.

The characteristics of the players' individual movement strategies and observed preparatory heights were reflected in the striking velocities. The most influential condition on the movement patterns was the dynamic level, resulting in higher striking velocities at higher dynamic levels. When comparing the striking surfaces, the players tended to increase striking velocity when playing on the soft surface, and to decrease striking velocity for the hard surface. Two of the players, (S2 and S4) also displayed a slight decrease in striking velocity for the stroke preceding the accent. The explanation can be found in the preparatory movements. Both these players initiated the preparation for the accented stroke by moving the hand and wrist upwards, letting the stick follow (see Figure 1). To reach the height

from which the accented stroke is initiated in time, the hand starts the upward movement even before the preceding hit. The upward direction of the hand and the stick makes the tip of the stick strike the surface much lighter compared to a normal stroke.

The main difference between the playing styles of the drummers was the emphasis on the accented stroke as compared to the unaccented stroke. This difference was noted both for preparatory heights and striking velocities. It is likely that the players' backgrounds could explain part of the differences in their emphasis on the accented stroke and movement strategies. Player S1 and S3 are mainly active in the symphonic and military orchestral traditions, while S2 and S4 mainly play the drum set in the Afro-American music tradition. In orchestral playing, an accent, although emphasized, should not be over emphasized. In contrast, many genres using drum set playing often encourages large dynamic differences between accented and unaccented beats. In fact, unaccented notes are sometimes played as soft as possible; "ghost notes".

For all players there was a clear tendency for lengthening every fourth interval, the one beginning with the accented stroke. It is worth noting that it is not the interval involving the preparatory movement before the accented stroke that is prolonged, but the following interval. Although the preparatory movements used for the accented strokes are larger (sometimes much larger, such as when player S2 involved the whole arm), the accented stroke arrives on time. It is the following stroke (the first stroke in the next measure) that is delayed. The lengthening not only emphasizes the accent, but also allows the oscillations of the rebounding stick to decay.

PERCEPTION OF MOVEMENT GESTURES

In the previous section we have seen examples of the movement strategies that musicians acquire to control also very simple elements in music. On a higher hierarchical level these "automated" control gestures are modified in accordance with the musical intention the performer wish to express. This higher layer concerned with musical expression was investigated through two studies on the perception of expressive intentions in performances on the marimba, bassoon and saxophone (Dahl and Friberg, 2004 and in press). The objective was to explore to what extent the movements conveyed the intended emotions and what movement cues observers used to discriminate between expressions.

Experiment I: Ratings of marimba performances

In Experiment I, a professional percussionist was video recorded when performing a short piece of music with the intentions Sadness, Anger, Happiness and Fear. From the video recordings, stimuli clips were generated showing different parts of the player in four *viewing conditions*: *full* (showing the full image), *no-hands* (the player's hands not visible), *torso* (the player's hands and head not visible), and *head* (only the player's head visible).

The clips were shown without sound, and edited so that facial expressions would not be visible. Twenty subjects watched the videos 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: slow - fast
Fluency: jerky - smooth
Regularity: irregular – regular

The ratings for the emotions showed that the intentions Sadness, Happiness, and Anger were successfully conveyed, while Fear was not. The intention receiving the highest number of correct identifications was Sadness. There was some confusion between the intentions Anger and Happiness, Anger rated as Happiness and vice versa, suggesting some features in common among these two emotional expressions.

The influence of viewing condition on the observers' ability to recognize the intentions was surprisingly small. However, the head of the player seemed to play a key role for some intentions. In the cases where there was an effect of viewing condition, the main differences in ratings were related to either the head or the torso conditions. For the Sad intention, all the conditions where the head was visible were rated high (means from 4.3 to 4.6), while the torso received much lower ratings (mean 3.1). For this intention 'head' was actually identified as Sad to a higher degree than was the 'full' condition. One explanation for this surprising observation could be that the actual playing movements of the hands and arms had a distracting effect on the movements of the head. Some support for this can be found in an earlier study. Davidson (1994) measured the head and hand movements of a pianist performing the same piece as "deadpan", "projected", and "exaggerated". She found that the extent of head movements differed significantly between performance conditions, while the extent of the hand movements was about the same.

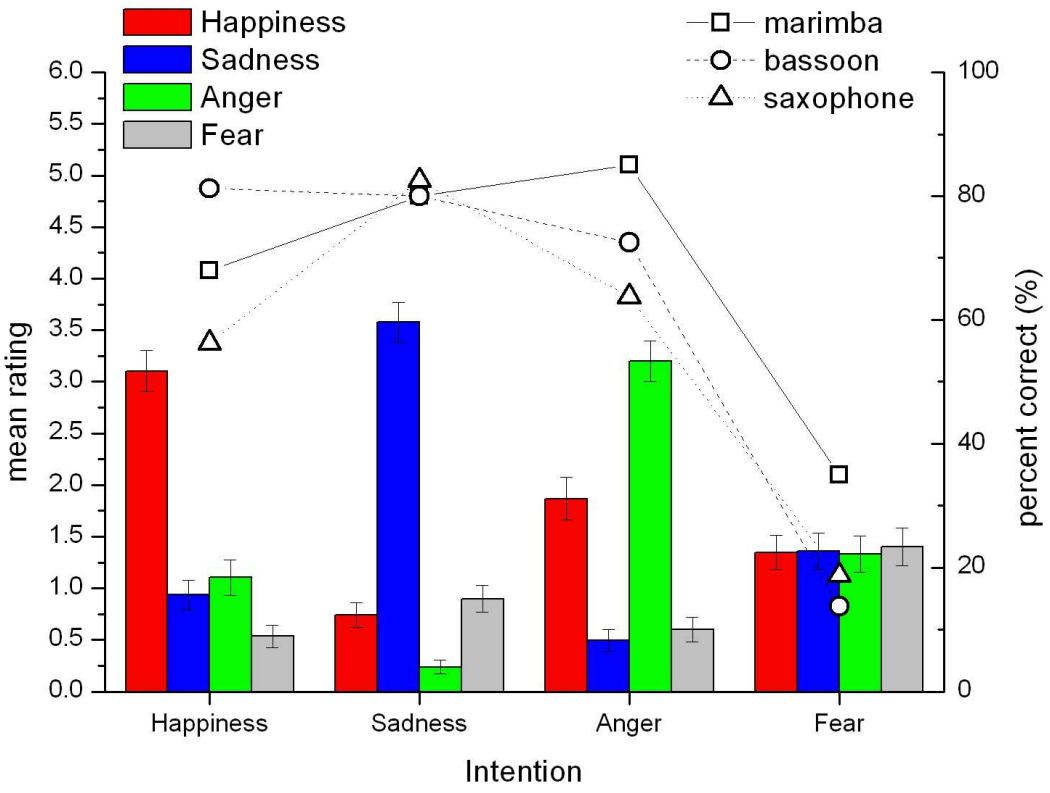
The ratings for movement cues showed that for the three recognized emotions Sadness, Anger, and Happiness, observers used movement cues similar to audio cues found for music performances. Anger was characterized by large, fast, uneven, and jerky movements; Happiness with large and somewhat fast movements, and Sadness with small, slow, even, and smooth movements.

Experiment II: Ratings of woodwind performances

To further explore the use of movement cues in the communication of specific emotions, and to investigate issues related to different types of instruments, a second experiment was carried out. In Experiment II observers rated video clips of a bassoon player and a saxophonist, performing four musical excerpts with the emotional intentions Happiness, Sadness, Anger, and Fear. No viewing conditions were generated for these clips. Twenty subjects rated the emotional content of the four emotions as well as the movement qualities rated in Experiment I, Amount, Speed, Fluency, and Regularity.

The emotion ratings of the woodwind performances confirmed the results of Experiment I. Happiness, Sadness, and Anger was all well communicated to the observers, while Fear was not. There was less confusion between the intentions Happiness and Anger. Compared to Experiment I, the movement ratings revealed fewer cues important for each emotion. Sadness was characterized by slow Speed and even Fluency, and Anger was characterized by jerky Fluency. Happiness was mainly characterized by fast Speed for the bassoon player, but large Amount for the saxophonist. The fact that the movement cues for Happiness were not overlapping with those for Anger, explains the lack of confusion between these two intentions compared to Experiment I.

Figure 2: Mean ratings for the four rated emotions averaged across subjects and cases in the two experiments. The bars show the mean ratings for each intended emotion, averaged across all subjects and rated video clips. The ratings show that Happiness, Sadness, and Anger all were conveyed to the observers while the communication of Fear failed. The lines show the per cent correct identified intentions for each of the three performers, bassoon player (circles), saxophonist (triangles), and the marimba player shown in the full viewing condition (squares). The transformation into per cent correct was strict. A 'correct' response was obtained only when the original intention received higher ratings than the other rated emotions. Happiness, Sadness, and Anger obtain between 56 and 85 per cent correct identifications, while the results for Fear are much lower, between 14 and 35 per cent.



The combined results for the emotion ratings for Experiment I and II are shown in Figure 2. The bars in the figure show the mean ratings for all the rated emotions and intentions, averaged across all three performers and all subjects and cases. As shown by the figure the three intentions Happiness, Sadness, and Anger all receive a large portion of ratings corresponding to the intention. The ratings for the intention Fear, however, are evenly spread between the four rated emotions. Also in the figure the percent correct for each intended emotion and player can be seen: bassoon player (circles), saxophonist (triangles), and the marimba player as shown in the full condition (squares). The values give the proportion of ratings for the intended emotion receiving the highest rating. While the values for Happiness, Sadness and Anger range between 56 and 85 per cent correct, the values for Fear are much lower, for the wind players even below chance level (25 per cent), indicating that the communication of Fear failed.

Table 1: Comparison between the most prominent movement cues found for the three performers and the corresponding audio cues (in *italics*); *sound level, tempo, articulation,* and *tempo variability* (selected from Juslin, 2001).

	amount ***sound level***	**speed** ***tempo***	**fluency** ***articulation***	**regularity** ***tempo variability***
Happiness	large *high*	fast *fast*	 *staccato*	 *small*
Sadness	small *low*	slow *slow*	smooth *legato*	regular *final ritard*
Anger	large *high*	fast *fast*	jerky *staccato*	irregular *small*
Fear	small *low*	 *fast*	jerky *staccato*	irregular *large*

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. Comparisons between the movement cues from the two experiments and the corresponding audio cues for music performance (selected from Juslin 2001) show several similarities (see Table 1).

DISCUSSION

The work reported here has exemplified both how performers use individual playing techniques to control low level features in the music, and what types of movement cues observers use to distinguish between specific expressive intentions. The control gestures for an individual player tend to be consistently used, although they vary considerably between players (see also Baader et. al., 2005, Engel et. al. 1997).

The player's hand and arm movements are primarily occupied with the production of notes. However, overlaid on the control gestures are overall movement cues that allow observers to extract information concerning the expressive intention. Being superimposed on the control gestures, these communicative gestures have to either *coincide* with the control gestures, or appear in parts of the body not to interfere with the actual playing movements.

As long as the same notes are played the control gestures only vary marginally between different performances. A player does not *have* to move in different ways in performances with different intentions – then why do it? A tempting answer is that these movements facilitate the interpretation of different emotions in some way. A possible support for this hypothesis is that the movement cues, as rated by the observers in the viewing test, indeed show strong similarities to the acoustic cues for the same emotions in music performances. Another, more straightforward, explanation is that the players tried to enter the intention they intended to convey and moved accordingly. Despite the differences in movement patterns between the players, the similarities dominate. Musicians seem to share a common body language which enables them to convey specific emotional intentions. 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.

From a control gesture point of view it is also possible that the performance on some instruments somehow benefits from overall body movements. Players of percussion and piano have limited influence of the tone once it is initiated. The time of direct control over the sound is much shorter than for, for instance, wind instruments or strings. Particularly, the onset times for notes differ substantially between these instruments. The well-defined onsets for piano and percussion result in high demands on the precision of the control gestures. The player has essentially no opportunities to adjust anything once the note is initiated, except for the decay. It is plausible that

playing of these short-contacts, “ballistic” instruments is easier to learn and control by using large and “gestural” playing movements as the large body sway in pianist performances reported by Davidson (1994).

REFERENCES

Baader, A. P., Kazennikov, O., and Wiesendanger, M. (2005). Coordination of bowing and fingering in violin playing. *Cognitive Brain Research*, 23:436–443.

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.

Dahl, S. and Friberg, A. (in press) Visual perception of expressiveness in musicians’ body movements. Manuscript accepted for publication in *Music Perception*

Dahl, S. and Friberg, A. (2004). Expressiveness of musician’s body movements in performances on marimba. In Camurri, A. and Volpe, G., (eds), *Gesture-based Communication in Human-Computer Interaction*, Volume 2915 of Lecture Notes in Artificial Intelligence, pp. 479–486. Springer Verlag.

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.

Engel, K. C., Flanders, M., and F., S. J. (1997). Anticipatory and sequential motor control in piano playing. *Experimental Brain Research*, 113:189–199.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and theoretical framework. In Juslin, P. and Sloboda, J. A., (eds), *Music and Emotion*, pp.309–337. 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, pp. 27 - 44. Kluwer Academic Publishers.

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.