Multimodal Analysis of Piano Performances Portraying Different Emotions

Jensen, Kristoffer; Frimodt-Møller, Søren

Published in:
From Sounds to Music and Emotions

DOI (link to publication from Publisher):
10.1007/978-3-642-41248-6_27

Publication date:
2013

Document Version
Early version, also known as pre-print

Multimodal Analysis of Piano Performances Portraying Different Emotions

Kristoffer Jensen and Søren R. Frimodt-Møller

Aalborg University Esbjerg, Denmark
krist@create.aau.dk, soren@frimodt-moller.dk

Abstract. This paper discusses the role of gestural vs auditive components of a piano performance when the performer is prompted to portray a specific emotion. Pianist William Westney was asked to perform a short passage from a specific piece of music 6 times, 3 times without making any deliberate changes, and 3 times where the music was intended to portray the emotions *happy*, *sad* and *angry*, respectively. Motion-capture data from all of the performances was recorded alongside the audio. We analyze differences in the data for the different emotions, both with respect to the size and shape of the pianist's movements and with respect to the sonic qualities of the performances. We discuss probable explanations of these differences. Although differences are found in both the gestural and auditive components of the performance, we argue that the gestural components are of particular importance to the performer's shaping of a musical expression.

Keywords: Music Informatics, Music Performance, Aesthetics, Gesture, Motion Capture.

1 Introduction

Several studies have been concerned with how music communicates emotions. According to Juslin [1], music elicits emotional responses in listeners, and performers are able to communicate anger, sadness, happiness and fear to listeners through tempo, sound level, frequency spectrum, articulation and articulation variability. It has also been shown [2] that the gestures of musicians communicate emotions efficiently. The current study proposes to compare the auditive aspects of a music performance with the gestural, analyzing how both vary across different music performances intended to communicate different emotions.

While the present study concerns piano performance, the movements described here have a lot in common with the movements of performers playing percussive instruments and to a large extent with the movements of musicians in general. Gestures are by definition movements that are expressive, but the way they are expressive varies. In relation to music, some gestures, which we will denote as *expressive gestures*, express something in addition to the produced musical sounds, they may emphasize certain moods or add a dramatic pathos to the performance. Gestures in a performance can, however, also be *effective* in the sense defined by Wanderley [3]:

M. Aramaki et al. (Eds.): CMMR 2012, LNCS 7900, pp. 469–479, 2013.

An effective gesture participates in defining the sound. (In this sense, one can say that it is not an expressive movement in itself, but the movement participates in producing music, which is, strictly speaking, a means of expression) As an example of effective gestures, Dahl [4] has investigated the preparatory gestures of drummers and showed that the drummers move the hand and the tip of the drumstick in a fishtail movement in order to gain the necessary height to perform certain notes. Many performance gestures, e.g. the movements of the pianists' arms and hands, are at the same time expressive and effective in the sense that they are both practically needed for the performance of certain notes and give the impression of being infused with a certain energy or emotion. Other gestures, such as head and facial ones may, however, primarily be regarded as expressive - they are apparently mainly part of a visual communication process with the audience. In the work of Davidson [5] (as well as the related work she provides a review of), facial gestures are argued to be an 'added value' on the emotions expressed by the pianist's entire torso, which, however, also seems to trace the phrase structure, dynamics and rhythm of the piece played. In this study, we consider gestures that may be both expressive and effective, but also comment on certain gestures that are arguably only expressive.

In February 2010, the Nordic Network for the Integration of Music Informatics, Performance and Aesthetics[1] held a workshop in Oslo, involving a session at the fourMs laboratory (Department of Musicology, University of Oslo) hosted by Professor Rolf Inge Godøy and postdoctoral researcher Alexander Refsum Jensenius. During this session, Pianist William Westney was prompted by the workshop participants (including the authors of this paper) to perform a short passage from "That Old Black Magic" (written by Harold Arlen, arranged for the piano by Cy Walter[2]) 6 times, 3 times without making any deliberate changes (denoted *normal* in the following), and 3 times where the music was intended to portray the emotions *happy*, *sad* and *angry*, respectively[3]. The performances were made on a Yamaha Disklavier. Motion-capture data from all of the performances was recorded alongside the audio. In this paper, we analyze differences in the data for the different emotions, both with respect to the pianist's movements, with the hand movements as an example, and with respect to the sonic qualities of the performances. When relevant, the emotions chosen are classified according to a two dimensional model [6], with arousal (the level of energy in the emotion) and valence (whether the emotion is positive or negative).

The motion capture equipment was a Qualisys motion capture system consisting of a nine-camera Oqus 300 system and a 200 fps grayscale Point Grey camera. Data was streamed in real time through OSC-protocol via UDP/IP to MAX/MSP/Jitter software and synchronized through a MOTU timepiece allowing synchronous playback of

---

[1] See `www.nnimipa.org` for more information on this research network.

[2] Sheet music for this passage can be found here: `http://www.cywalter.com/archives/SheetMusic/ThatOldBlackMagic/HTMLs/Page2.htm` (retrieved by April 29, 2013).

[3] Videos, alongside MoCap visuals of the three emotion-laden performances can be found under part 3 of [11], `http://nnimipa.org/JWG.html`

analog, motion, video and audio data [7,8]. MoCap data of 23 different points[4] on Westney's head, torso, arms and lower body are available.

2 Audio Analysis

For the analysis of the audio, 4 different features were estimated from the sound files: the sound pressure level (SPL), according to the ISO-226 standard (2003); the dynamics, calculated as 90% of the max volume minus the median, divided by the median; the spectral centroid (SC) - a relative measure of the strength of the higher partials (related to brightness); and the sensory dissonance (SD), calculated as the sum of all overtone pairs that cause fluctuations, weighted with the amplitude of the overtones [9]. The latter is related to musical tension.

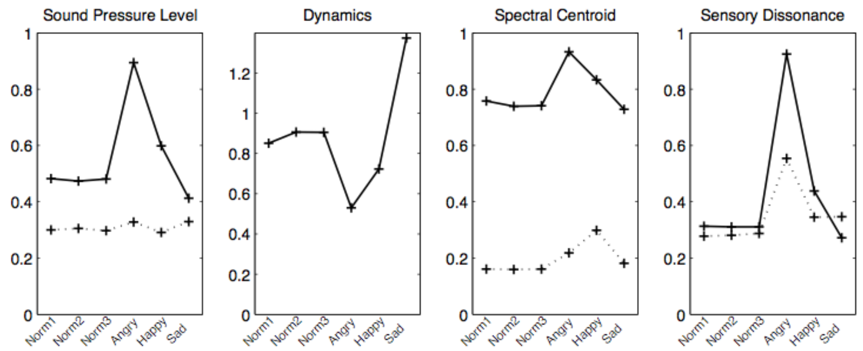

Fig. 1. Mean (solid) and standard deviation (dotted) of Sound Pressure Level (left), Dynamics, Spectral Centroid and Sensory Dissonance (right). Dynamics are calculated globally, so the value has no mean or standard deviation.

In Figure 1, the four features, as calculated from the six performances are shown. It can be seen that the three *normal* performances have very much the same values of these features, while the 'emotional' performances have varying values. In particular, the *angry* and, to a lesser degree, the *happy* performances have higher sensory dissonance, sound pressure level and spectral centroid, but less dynamics, while the *sad* performance has lower dissonance, loudness and brightness and more dynamics than the *normal* performances. The lower degree of dynamics in the *angry* and *happy* performances is obtained because very little of the audio has low loudness, thus restraining the dynamic range. Similarly, the *sad* performance has more dynamics, because even though most of the audio has low loudness, a few very loud notes occurred. In addition, the tempo of the six performances were estimated, showing that the *angry* and, to a lesser degree, the *happy* performances were played significantly faster (154 & 138 BPM respectively), and the *sad* performance slower (94 BPM) than the

[4] `http://fourms.wiki.ifi.uio.no/MoCap_marker_names`. Some of these are omitted in the current study.

normal performances (all three at 104 BPM). (All tempi given as measured at start of piece.)

3 Video Analysis of Hand Movements

As an example of how the movements of a performer can be analyzed, we consider the movement of two points on William Westney's body (out of the 23 that were mapped by the Motion Capture system), namely the right and left hand inwards markers (RHAI and LHAI), placed on upper part of the hand under the index finger of each hand. (The movement of Westney's head will be discussed in section 4.) The movements of these are shown in Figure 2 for the six performances. (The y-axis indicates the horizontal distance from the keyboard, whereas values on the x-axis indicate the horizontal position along the keyboard direction. The path of each graph shows how the position of the hands shifts over time.) The three *normal* performances have approximately the same gestures, which is shown below. Some of these are simply related to the performance of specific notes, others are effective, e.g. the movement of one hand away from the keyboard in order to make room for the other hand.

For the emotion-laden performances, the movements seem (in Figure 2) to be larger for *angry* and *happy* compared to the three *normal* performances, but smaller for *sad*, supporting the idea that *angry* and *happy* prompt larger gestures in the performance than *sad* does. The *angry* and *happy* performances also seem to have more preparatory movements (made before starting the performance, at the vicinity of 'start' in Figure 2) at the start and end of the performances, again supporting the idea of *angry* and *happy* as more lively gestures than *sad*. Some of the differences (such as the lack of the arc in the *normal* performances) may be attributed to the performer changing his approach. The more detailed analysis given in the sections below reveals interesting differences between the movements made during the different emotions. When reading the following sections, it is, however, important to keep in mind that the standard deviation is generally too large, and varies too much between different conditions, to allow for any statistical significance of the differences between the emotions. More data, i.e. more musicians and more pieces of music would be necessary in order to perform a study with statistically significant differences.

3.1 Height, Width and Depth of Movements

In order to measure the relative size of the hand movements in the different performances, we detect the positions in each of the three spatial dimensions where the performer changes direction. For all positions, we measure the distance between the lower and higher points. Finally we calculate the mean of these distances for each of the three dimensions.

Figure 3 shows the mean movement distances along the x- y- and z-axes for each of the six performances.

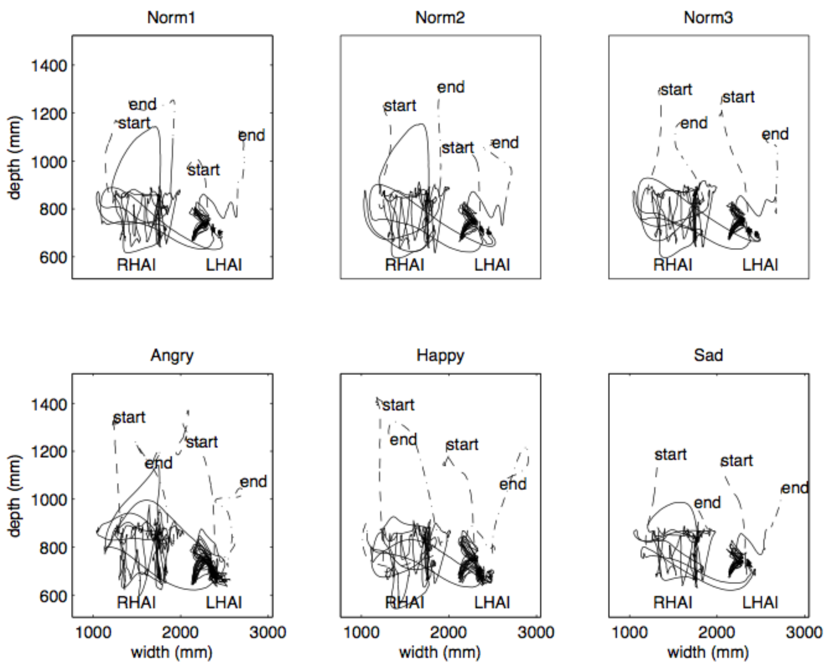

Fig. 2. Movement of RHAI and LHAI markers

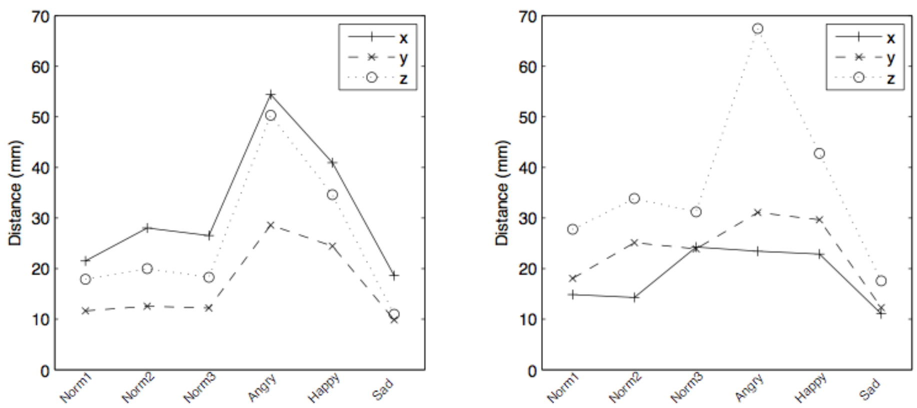

Fig. 3. Mean movement distances along the x- y- and z-axes for each of the six performances. The left figure shows the values for the LHAI markers, the right figure for the RHAI markers. The standard deviation (not shown) is generally as high as or higher than the means.

In all of the performances, the right hand displays larger movements than the left if measured along the vertical dimension (the y-axis), whereas the left hand in general has ‘wider’ movements, i.e. the largest movements when measured along the x-axis. Both of these facts can be explained in relation to the piece of music being played:

The left hand mainly plays accompanying chords, whereas the right hand mainly plays the melody. The left hand moves to the right of the right hand at one passage, while the right hand typically stays within the same octave on the keyboard. The left hand stays closer to the keyboard to make the transitions between chords easier, whereas the right hand has more freedom to move upwards, away from the keys, in the small pauses between notes. For the same reason, the values for movement along the z-axis (height), are also slightly smaller for the left hand than for the right.

Movements along the x-axis are in general tied to which notes are being played. It is, however, worth noting that the values for movements in the left hand along the x-axis are in fact higher for *angry* and *happy* than for the other performances. In fact, when the distance values for one dimension increase, the other two dimensions follow suit. This may indicate that the pianist is not able to isolate the expressive energy he infuses his gestures with to just one spatial dimension - or even one limb. Put differently, if he wants to move, say, his right hand more violently up and down, this character spreads to all of his movements, so that left hand movements from side to side also become more intense.

As seen in Figure 3, *angry* has larger movement distances along the y-axis than the other performances (for both right and left hand), indicating that the pianist in general plays the notes more violently when trying to capture this emotion. This is caused (as found in data not shown here) by the performer starting from a greater height before each attack for *angry* than in the other performances. *Angry* and *happy* both display larger movement distances in all three dimensions for the left hand and along the y- and z-axes for the right hand, whereas *sad* has lower distance values for all of the three dimensions than any of the other performances. This supports the idea that *angry* and *happy* involve a *high arousal level*, meaning that they are types of emotions that have a lot of energy in them, whereas *sad* prompts a more timid and calm attitude of the performer. This mirrors the finding in Figure 2, that *angry* and *happy* has higher SPL, brightness and sensory dissonance, and *sad* has lower, when compared to the *normal* performances.

3.2 Speed and Curvature

The size of the hand movements is, however, only one parameter of a performer's gestures. The actual shape of the gestures is just as important, as well as the speed with which they are carried out, e.g. the degree of abruptness with which the hands move.

'Pointy' edges in the graphs in Figure 2 suggest that the pianist plays more staccato, that is, moves his hand away fast after the attack, whereas softer edges indicate a gentler movement. In order to assess this trait of the motion capture data, the Euclidean speed and the curvature (how bent the gesture is; calculated as the length of the vector cross product of the first and second time derivative of the positions divided by the length of the time derivative to the power of 3) has been calculated for the different markers (low curvature corresponds to a high degree of 'pointiness' in the graph). The mean and standard deviation of the speed and curvature is shown in Figure 4. With regards to the speed, it is clear that the emotions *angry* and *happy* have higher speed in comparison with the *normal* performances, while the *sad* performance has lower speed. The curvature values are markedly lower for the *angry* and *happy* performances than for the other performances. While there is much noise in the curvature values (c.f. the standard deviations), intuitively, the curvature values seem reciprocal to the speed. This is related to the 1/3 power law [10], which states that, for a specific gesture, the angular speed is equal to a constant (k) multiplied with the radius to the power of one third. As the curvature is the reciprocal of the radius, it is normal to expect the curvature to be inversely proportional to the speed. The constant (k) is calculated as the mean of the speed multiplied by the curvature to the power of one-third and found to be proportional to the arousal value of the emotion, as the *high arousal*-emotions *angry* and *happy* [6] are found to have a constant (k) higher than in the *normal* performances while the *low arousal*-emotion *sad* has a constant (k) below the one in the *normal* performances. It is interesting to observe how the speed and the one-third power law constant (k) (not shown in figure) are systematically lower for the left hand. This is probably related to the fact that the left hand has less general movement, but whether this is further related to the music performed here, or a potential right-handedness of the performer is a matter for further studies beyond the reach of this work. In short, when Westney plays *happy* or *angry*, the speed is higher, and the curvature is lower. Speed is then correlated to the audio features SPL, brightness and sensory dissonance, i.e. when the speed is higher then the audio features is also higher.

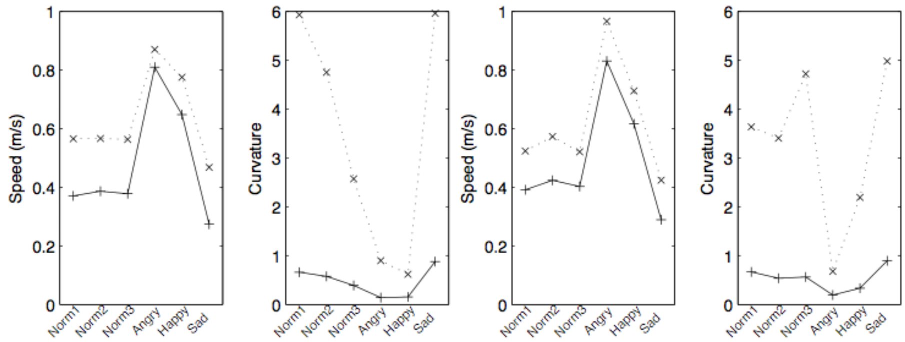

Fig. 4. Mean values (solid) and standard deviations (dotted) for Euclidian speed and curvature. Left hand (left) and right hand (right)

4 Exclusively Expressive Gestures

So far, we have only considered how differences with respect to how the pianist shapes his hand gestures in the performance coincide with his attempt to infuse the performances with particular emotions. Gestures involving other parts of the body might, however, also be important parts of the musician's expressive means. In the following we look first at the role of the pianist's head in his performance gestures and postures, and then proceed to consider the role of his facial gestures. These gestures have in common that they do not seem to be effective, although in theory part of the movement of the pianist's head might intuitively follow other movements in his torso which are caused by effective gestures in the arms and hands.

4.1 Head Posture

When studying the movements of the trackable points on William Westney's head, we found that while the movement of Westney's head were not significantly large or different enough to warrant any conclusions with respect to their expressivity, the *position* of his head was in fact quite different from performance to performance. More specifically, we found interesting differences in how high Westney holds his head over the piano in the different performances, and also in the slope of his head, that is, how it is inclined. The results for the 6 video recordings are showed in Figure 5 below.

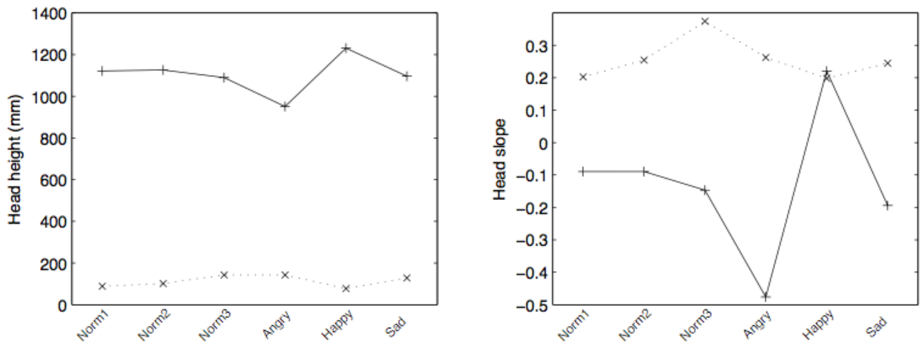

Fig. 5. Head height (left) and slope (right) for the six performances. Mean values are indicated with solid '+' and standard deviation with dotted 'x'.

The head height is measured using four head markers (left and right and front and back), and calculated as the mean of the front left and right head height, and the slope is calculated by dividing the differences between the front and back height and depth values. It is clear that the head is lower and more inclined forward in the *angry* performance, and higher and more inclined backwards in the *happy* performance. These results points to a significant correlation between the valence of the intended infused emotions and the position of the head: Positive valence corresponds to a positively inclined head held high, while negative valence coincides with a negatively inclined head held low. Camurri et al [12], in an observation study of a pianist's gestures and their role in the interaction with the audience, show that the pianist moves his head more in a markedly expressive performance than in a *normal* performance. Our result above expands on this observation in the sense that we are able to distinguish between different kinds of expressivity and their relation to different head postures. Camurri et al also hypothesize that the back contracts along with heightened arousal or build-up of tension while opening when this tension is released. Our study (deduced from head position in Figure 5, and confirmed by C7 marker depth, not shown) shows that this process goes in both directions dependent on the valence: the body leans forward and the head lowers with negative valence and the body leans backwards and the head rises with positive valence. This differs from the audio analysis, in which *angry* and *happy* generally had larger feature values, and thus it does not seem to be related to the sound production, at least not to the features estimated from the audio.

4.2 Facial Gestures

In ensemble performances, facial gestures are important parts of the interaction between the players, e.g. in a string quartet, in the sense that they tend to communicate very specific information (e.g. "it is your turn to play now" or "play softer"). Isabella Poggi [13] has described how a conductor's facial gestures during symphony orchestra performances even constitute an almost unambiguous sign system (in relation to the orchestra musicians). In the case of one person performing, i.e. a solo performance, the role of facial gestures seems less well-defined. Thorough observation of the 6 video recordings of William Westney showed that there was little connection between the intended emotion to be conveyed and the facial expression of the pianist. In the three *normal* performances, one sees an inkling of a smile on the pianist's face 32 seconds into the first video, and in the other two videos, a solemn look at the start and again at about 15-18 seconds into the performance. In the *angry* performance, Westney looks down so much that it is not possible to see his facial expressions properly. In the *happy* video, the performer displays a solemn smile throughout, while the *sad* performance has the performer looking downwards a lot. Other than these subtle differences, the pianist keeps a fairly expressionless face throughout the performances.

Our tentative conclusion regarding facial gestures in Westney's solo performance is that they do not play any important part in his communication with the audience, given that they do not seem to display any of the emotions that Westney is trying to portray in the situation. All that can be observed is a solemn smile that is not properly speaking *happy*.

5 Conclusions and Further Perspectives

We have shown specific differences between the *normal* and the *angry*, *happy*, and *sad* performances of one piece of music played by a pianist with respect to sound level, spectral centroid, dynamics and sensory dissonance in the audio, and for the hand movement sizes, speed and curvature. We have also shown that the pianist's head posture seems to follow changes in the intended infused emotions. While the size, speed and curvature are directly correlated with the audio features, the head slope is not, as it accounts for valence changes that are not found in the audio features.

We also looked for differences across the different performances with respect to the pianist's facial gestures, and found that the pianist was smiling in the *happy* performance. To sum up, while there are also marked differences between the audio from the 6 performances, gesture (as observed for the hands and head) seems to be a very important component for the pianist, when prompted to shape his expression according to a specific emotion.

Given that the study only uses data from one pianist and one piece of music, the findings only give a general idea with respect to tendencies in the expressive means of a performer. We have also not considered whether the gestures of the performer actually give the audience an experience of the different emotions the performer intends to portray. Possible expansions of this study thus includes repeated experiments with more pianists, different pieces of music, and data from the audience (e.g. collected via questionnaires) with respect to how they experience the performance.

With respect to the shape of gestures, a further analysis of the movements of the other 21 points on Westney's body might yield more nuanced results. Nonetheless, this study shows interesting observations on the differences in both audio and gestures for different emotional expressions, indicating how performances rely on both modalities when conveying emotions.

References

1. Juslin, P.: 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 (2000)
2. Dahl, S., Friberg, A.: Visual Perception of Expressiveness in Musicians' Body Movements. Music Perception: An Interdisciplinary Journal 24(5), 433–454 (2007)
3. Wanderley, M.M.: Non-obvious performer gestures in instrumental music. In: Braffort, A., Gibet, S., Teil, D., Gherbi, R., Richardson, J. (eds.) GW 1999. LNCS (LNAI), vol. 1739, pp. 37–48. Springer, Heidelberg (2000)
4. Dahl, S.: Measurements of the motion of the hand and drumstick in a drumming sequence with interleaved accented strokes: A pilot study. In: Report Speech, Music and Hearing, Quarterly Progress and Status Report, 4/1997, pp. 1–6. Royal Institute of Technology, Stockholm (1997)
5. Davidson, J.W.: Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies. Psychology of Music 40(5), 595–633 (2012)
6. Russell, J.A.: A Circumplex Model of Affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
7. Morander, C.: FourMs - Music, Mind, Motion, Machines: How Can Knowledge About Movement & Sound Create Better Technologies? Qualisys Newsletter (1), 4 (2010)
8. Jensenius, A.R., Glette, K.H., Godøy, R.I., Høvin, M.E., Nymoen, K., van Dorp Skogstad, A., Tørresen, J.: FourMs, University of Oslo - Lab Report. In: Rowe, R., Samaras, D. (eds.) Proceedings of the International Computer Music Conference, New York, June 1-5 (2010)
9. Sethares, W.A.: Local Consonance and the Relationship between Timbre and Scale. Journal of the Acoustical Society of America 94(3), 1218–1228 (2003)
10. Lacquaniti, F., Terzuolo, C., Viviani, P.: Global Metric Properties and Preparatory Processes in Drawing Movements. In: Kornblum, S., Requin, J. (eds.) Preparatory Processes in Drawing Movements, pp. 357–370. Lawrence Erlbaum, Hillsdale (1984)

11. Jensenius, A.R., Nymoen, K., Grund, C.M., Westney, W., Skogstad, S.A.: Video Suite – in Three Movements: Jensenius-Westney-Grund on Motion-capture, Music and Meaning. Multimodal webpage presentation (2010), http://www.nnimipa.org/JWG.html
12. Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., Volpe, G.: Multimodal Analysis of Expressive Gesture in Music and Dance Performances. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 20–39. Springer, Heidelberg (2004)
13. Poggi, I.: The Lexicon of the Conductor's Face. In: McKevitt, P., Nuallin, S., Mulvihil, C. (eds.) Language, Vision and Music: Selected papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing, Galway, Ireland 1999, pp. 271–284. John Benjamins Publishing Company, Philadelphia (2002)