Automated Analysis of Body Movement in Emotionally Expressive Piano Performances

Ginevra Castellano
InfoMus Lab, University of Genoa, Genoa, Italy

Marcello Mortillaro
Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Antonio Camurri and Gualtiero Volpe
InfoMus Lab, University of Genoa, Genoa, Italy

Klaus Scherer
Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

EMOTIONAL EXPRESSION IN MUSIC PERFORMANCE includes important cues arising from the body movement of the musician. This movement is related to both the musical score execution and the emotional intention conveyed. In this experiment, a pianist was asked to play the same excerpt with different emotionally expressive intentions. The aim was to verify whether different expressions could be distinguished based on movement by trying to determine which motion cues were most emotion-sensitive. Analyses were performed via an automated system capable of detecting the temporal profiles of two motion cues: the quantity of motion of the upper body and the velocity of head movements. Results showed that both were sensitive to emotional expression, especially the velocity of head movements. Further, some features conveying information about movement temporal dynamics varied among expressive conditions allowing emotion discrimination. These results are in line with recent theories that underlie the dynamic nature of emotional expression.

Received November 18, 2006, accepted May 30, 2008.

Key words: Emotional expression, bodily movements, piano performance, automated analysis, emotional discrimination

THE RELATIONSHIP BETWEEN MUSIC and emotions is well acknowledged (Cooke, 1959; Langer, 1957) and has been investigated in terms of expression, recognition, and induction (Juslin & Laukka, 2004; Scherer, 2004; Sloboda & Juslin, 2001). Emotional effects of music can be produced through structural characteristics (Clarke, 1988), performer's ability, personal associations of listeners and the context where performance and listening take place (Gabrielsson & Juslin, 2003; Juslin & Laukka, 2004). All these aspects play a role in inferential and induction processes within listeners (Scherer & Zentner, 2001). This view is in line with a Brunswikian perspective (Brunswik, 1956), where a comprehensive account of emotion communication should consider the entire path from sender to receiver. On the sender side, transient states (like emotions) are expressed in an organism's appearance and behavior by means of cues, which can be objectively measured. These cues—called distal because they are remote from the receiver—correlate with the sender's state and, according to Brunswik's terminology (1956), the degree of correlation constitutes the ecological validity of the cues. On the receiver side, distal cues are the object of a perception process. Perceived cues—called proximal because they are close to the receiver—constitute the basis on which the receiver makes inferences to attribute states to the sender. This model emphasizes the fact that objectively measured distal cues are not necessarily equal to perceived proximal cues. Proximal cues are based on distal cues, but the latter may be modified by the transmission channel and by structural characteristics of perceptual systems. This implies that a comprehensive account of emotion communication requires the inclusion of both expression and perception (Scherer, 1978, 2003).

Studies on emotional expression in music have largely focused on acoustic measures. Reviews by Gabrielsson and colleagues claimed the existence of a quite high recognition rate of intended emotional expressions, for at least broad emotion categories or dimensions, proposing that specific configurations of musical factors correspond to definite emotions (Gabrielsson & Juslin, 2003;

Music Perception VOLUME 26, ISSUE 2, PP. 103–120, ISSN 0730-7829, ELECTRONIC ISSN 1533-8312 © 2008 BY THE REGENTS OF THE UNIVERSITY OF CALIFORNIA. ALL RIGHTS RESERVED. PLEASE DIRECT ALL REQUESTS FOR PERMISSION TO PHOTOCOPY OR REPRODUCE ARTICLE CONTENT THROUGH THE UNIVERSITY OF CALIFORNIA PRESS'S RIGHTS AND PERMISSIONS WEBSITE, HTTP://WWW.UCPRESSJOURNALS.COM/REPRINTINFO.ASP. DOI:10.1525/MP.2008.26.2.103

Gabrielsson & Lindström, 2001). These regularities in emotional expression and recognition from music have also been confirmed in studies concerning synthesis and production. De Poli and colleagues (Canazza, De Poli, Drioli, Rodà, & Vidolin, 2000; De Poli, Rodà, & Vidolin, 1998) applied an analysis-by-synthesis methodology for deriving a model able to synthesize an emotional tone starting from a neutral one. From perceptual tests, they obtained a Perceptual Parametric Space, mapping expressive intentions (e.g., dark, light) on a 2D space with axes related to kinetics (tempo and articulation) and energy (loudness). Similarly, Sundberg and colleagues developed a rule-based system for generating musical expressive performances, in terms of acoustic variations (Friberg, 1995; Sundberg, Friberg, & Frydén, 1991; Thompson, Sundberg, Friberg, & Frydén, 1989). Rules describe how musicians modify the nominal score execution depending on their expressive intentions and affect several aspects of the performance such as duration of tones, loudness, pitch, vibrato, crescendos and decrescendos, tempo, and articulation.

While many studies confirmed that emotions can be communicated by music with better-than-chance accuracy (Gabrielsson & Juslin, 1996; Juslin, 1997a, 1997b), performance studies were not able to show fixed configurations of cues across, and even within, performers (Juslin, 2000). The loose association of performance characteristics and emotional expression can be partly understood on the basis of the multimodal nature of human communication. Emotions are generally expressed through different channels at the same time, namely face, voice, speech, and body movements. Music performance involves the same channels of expression (Livingstone & Thompson, 2005). For example, it was shown by Di Carlo and Guaitella (2004) that emotions in music—in live performances and audiovisual recordings—are communicated not only through sound but also through facial expressions and other body movements. In a wider perspective—considering an ethnographic, empirical, and historical/cultural perspective—Thompson, Graham, and Russo (2005) claimed that visual information influences perception and experience of music. In a recent study Thompson, Russo, and Quinto (2008) showed that facial expressions of performers convey emotional content and influence the perception of emotion in listeners. They also demonstrated that the integration of visual information and auditory cues influences the emotional perception and interpretation of listeners automatically and pre-attentively.

Psychology has a long tradition of studies focusing on the relationships between emotion and movement qualities (Boone & Cunningham, 1998; De Meijer, 1989; Pollick, Paterson, Bruderlin, & Sanford, 2001; Scherer & Wallbott, 1985; Wallbott, 1998; Wallbott & Scherer, 1986). While some studies found evidence for characteristic body movements accompanying specific emotions (e.g., Boone & Cunningham, 1998), others argued that movements may be only indicative of the intensity of emotion, but not of its quality (Ekman & Friesen, 1974). Wallbott (1998) analyzed body movements and postures of actors portraying different emotions, elicited via a scenario approach. Results showed some distinctive patterns of movement and postural behavior associated with some of the emotions studied. While specific patterns did not emerge, several distinctive features that allow some degree of distinction between emotion categories were found. For example, lifting the shoulders seems to be typical for joy and hot anger while moving shoulders forward is frequent for disgust, as well as for despair and fear.

The influence of body movement on the perception of emotion in music performance has already been proposed in some studies (Dahl & Friberg, 2007; Timmers, Marolt, Camurri, & Volpe, 2006; Vines, Wanderley, Nuzzo, Levitin, & Krumhansl, 2004). In a recent and relevant contribution, Dahl and Friberg (2007) investigated how emotional intentions can be conveyed through musicians' movements. Namely, the role of the different body parts during an emotional expression was rated. It was found that head movement played an important role in the communication of emotional content. Similar results were presented previously by Davidson (1994), who found that the head was important for observers to discriminate between piano performances with different expressive intentions.

Whereas in the literature there are studies reporting the influence of movement in the perception of emotion in music, generalization of results is impaired by the lack of a shared coding scheme for movement and by the difficulty of replicating studies using only perceptual measures. Manual coding of movements mainly allows for a categorical or qualitative description of predefined tokens, while emotional expression in the body may be better described in terms of continuous variables, including temporal information. The detection of these features cannot be done manually, but instead, requires automated extraction software.

In human-computer interaction, there is increasing attention on automated analysis techniques aiming to extract and describe information related to the emotional state of individuals. In particular, some attempts were made towards the design of systems capable of analysing expressive body movements and automated emotion recognition.

Camurri, Lagerlöf, and Volpe (2003) classified expressive gestures in human full-body movement, namely in dance performances. They identified cues deemed important for emotion recognition and investigated how these cues could be tracked by automated recognition techniques. In particular, they showed that some motion cues—overall duration of time, contraction index, quantity of motion, and motion fluency—explained the differences in a choreography performed with four different emotional expressions—anger, fear, grief, and joy. On the basis of these motion cues authors defined an automated classifier able to distinguish among the four conditions.

Castellano, Villalba, and Camurri (2007) used similar motion cues, such as quantity of motion, speed and fluidity of movement, to infer acted emotions in order to investigate the role of movement expressivity versus shape in gestures. They proposed a method for the analysis of emotional behavior based on both a direct classification of a number of time series of motion cues and a model that provides indicators describing the dynamics of expressive motion cues. They found that the quantity of motion and contraction index of the upper body played a major role in discriminating between different emotions.

Bernhardt and Robinson (2007) proposed a framework for analyzing motion-captured knocking motions (e.g., in the sense of knocking on a door). They used statistical measures of movement qualities (e.g., the velocity and acceleration of the hand) computed over motion primitives that proved successful in the recognition of implicitly communicated affect.

Other studies showed that the performance of automated emotion recognition systems increases when different modalities are integrated. Balomenos and colleagues (Balomenos, Raouzaiou, Ioannou, Drosopoulos, Karpouzis, & Kollias, 2005) for example, proposed an approach in which gesture recognition was used jointly with facial expression analysis in a bimodal emotion recognition system. Further, Gunes and Piccardi (2007) fused facial expression and body movement information at different levels in order to conduct bimodal emotion recognition. They found that the accuracy of emotion recognition was improved when expressive body and face information was considered, in comparison to considering the face alone.

This paper focuses on video analysis of emotional expression in music performances. It addresses the dynamics of gestures used by music performers to communicate their emotional expressive intentions. From a Brunswikian perspective, the present research analyzes the distal cues of emotion expression in the body movements of a piano player.

An experiment was conducted where a pianist was asked to play a fragment of the same piece (an excerpt from the Sonata No. 4, Op. 102/1 for piano and cello from Ludwig van Beethoven) with different emotional expressive intentions. Following some of the suggestions of Scherer (Scherer, 2004; Scherer & Zentner, 2001) it was decided to use more subtle labels—referring to emotionally expressive modes—than traditional basic emotional categories: personally felt affect, sad, allegro, serene, overexpressive. Sad, allegro, and serene are quite traditional expressive modes in music performance, which can be easily understood by performers and applied to a music piece. In the personally felt affect condition the pianist was asked to play using the expressive mode she thought was the most appropriate to the piece. In the overexpressive condition, movement expressivity was particularly emphasized while a definite emotional characterization was absent.

In this paper an exploratory approach for the analysis of expressive movement in music performance is proposed, based on the dynamics of the pianist's gestures.[1]

In the proposed approach, expressive motion cues from gestures captured during five performances with different emotional expressions were extracted and analyzed. Expressive motion cues are descriptors providing information about how a gesture is performed, e.g., in an impulsive or in a smooth way, with body contraction or expansion, fluent or hesitant. Further, extraction and analysis were supported by statistical and computer engineering techniques.

In this study, the temporal profiles of two specific motion cues are considered: the quantity of motion of the upper body and the velocity of the head movements of the pianist. This choice was supported by results obtained in previous studies, where these two cues were found to be successful in conveying expressive information. Camurri et al. (2003) noticed that the quantity of motion is a relevant cue in recognizing emotion from full-body movement of dancers. Further, Dahl and Friberg (2007) showed that the movement of the head provides important information for identifying the emotional intention in marimba performances, while

[1] The general goal of the current study was to assess the feasibility of the newly developed approach. In relation to this task, which was exploratory in nature, an analysis of a single musician performing a music excerpt in different conditions seems to be sufficient.

Davidson (1994) found that head movement is important for observers to discriminate between pianist's performances played with different expressive intentions. On the basis of these studies, we decided to include these two motion cues and apply them to a different domain (in case of quantity of motion) or to replicate them with a different experimental design (velocity of the head movements). In addition, the current approach particularly emphasizes the temporal aspect of the cues, so reflecting the dynamics of the pianist's movements.

The first objective is to verify whether specific continuous movement measures, previously identified in studies on dance performances and acted emotions (Camurri et al., 2003; Castellano et al., 2007), could be also used in evaluating the emotional expression in music performances. It is expected that quantity of motion is sensitive to emotional expression in music performance. In addition, replications about the relevance of head movements are expected in terms of velocity.

The second objective concerns the discrimination of different emotional expressions on the basis of selected features, which convey information about the dynamics over time of quantity of motion and velocity of the head movements. As it is for emotional expression in face, body, and voice (Scherer, 2001) and for musical expression (Gabrielsson & Juslin, 2003), it is expected that time is a key element for disentangling emotional performance in music. In particular, it is supposed that, given the constraints of score playing, it is the gesture temporal dynamics that plays a role in the communication of emotional expressive content.

Thirdly, this study assesses the suitability of these features for emotion discrimination. Some other studies (Bernhardt & Robinson, 2007; Camurri et al., 2003; Castellano et al., 2007) found that it was possible to discriminate emotions on the basis of movements, but, to our knowledge, no specific findings for music performances are currently available.

Method

Materials

The experimental recordings were carried out in the repetition hall of the *Orchestre de la Suisse Romande* in Geneva. One professional concert musician, a pianist, performed an excerpt from Sonata No. 4, Op. 102/1 for piano and cello by Ludwig van Beethoven (Allegro vivace, measures 28-55) with five different emotional intentions.[2] Instead of using discrete emotion categories, following discussion with the performer, it was decided to use a number of emotion expression categories that are frequently used by musicians and that can be mapped onto dimension of valence and arousal. Through an imagery procedure the pianist was instructed to perform the excerpt in five modes, labelled with emotional or musical terms: personal, sad, allegro, serene, overexpressive. In order to obtain a performance to be considered as a baseline condition, the pianist was told to perform in a manner she thought was most appropriate to the affect in the respective passage and in the expressive manner she would adopt in performing the piece (*personal* condition). *Sad* refers to a low activation and negative valence condition; *allegro* refers to a medium-high activation condition, and it mainly refers to a tempo marking; *serene* was used as a positive emotion with a medium level of activation; *overexpressive* refers to an interpretation that exaggerates the musical performance without a definite emotional meaning and it was used as an extremely activated condition, neither positive nor negative. Labels were sequentially presented to the pianist who was asked if she was familiar with the quality associated with each condition and if she felt that she could produce the respective interpretative quality. In some cases, she repeated a performance to achieve the appropriate quality to her satisfaction.

Two video cameras (SONY DSR-PDX10) with constant shutter, manual gain, and focus at 25 fps were used to record the movement of the pianist. The pianist was asked to wear light clothes and dark panels were fixed behind her in order to allow, during the video analysis, the tracking of her body by separating it from the background. The pianist was video recorded from two sides—lateral and top views. A lateral view of the pianist was necessary for the computation of the quantity of

[2] In the original experimental set-up, a cello player also was included. These recordings were not considered within the present study because each instrument limits the player in a specific way. It is expected that players of different instruments communicate emotional expressive intentions using, consciously or unconsciously, different motion cues that might not be independent from technical movements. The kind of measures herein adopted thus should have been slightly modified to compare players of different instruments; this is also due to some technical difficulties (for example, the camera point of view should be changed). Therefore, the present study deals with the expressive characteristics of the movements of a single piano player and does not address the comparison between different instruments players, which could be a topic of future studies.

motion of her upper body. A top view was recorded for the assessment of the head and back movements.

Gesture Segmentation

An expert musician segmented the pianist's movement in all performances into gestures based on the musical score—so that each gesture corresponded to a musical phrase or sub-phrase. Further, the extremes of the gestures were adjusted considering the separation between two consecutive gestures given by a minimum of the quantity of motion (QoM) profile (see next paragraph for its mathematical definition) as reported in literature (Camurri et al., 2003; Kahol, Tripathi, Panchanathan, & Rikakis, 2003). Hereafter, we will refer to *gesture* as being the pianist's movement related to a musical phrase or sub-phrase, characterized by a positive initial slope and a negative final slope of the QoM profile. Fifteen gestures were located in the musical score, yielding a total of 75 gestures (15 gestures × 5 conditions). QoM and velocity temporal profiles were calculated in each gesture.

Measures

In a cross-disciplinary perspective, research on expressive cues describing emotional aspects of human motion and gesture can be built on several bases, ranging from biomechanics, over psychology to performing arts theories. The present research considers theories from choreography like Rudolf Laban's Theory of Effort (Laban, 1963; Laban & Laurence, 1947), work by psychologists on nonverbal communication (e.g., Argyle, 1980), and on expressive cues in human full-body movement (Boone & Cunningham, 1998; Wallbott, 1998). Based on the above mentioned theories and on previously reported findings (Camurri et al., 2003; Castellano et al., 2007; Dahl & Friberg, 2007; Davidson, 1994), two motion cues were defined, i.e., descriptors providing information about movements and gestures. Techniques (computational models, algorithms) for extracting motion cues from visual data were developed as well as software modules implementing these techniques.

In order to analyze the movement of the pianist, motion cues were automatically extracted from the video recordings. The layered approach proposed by Camurri (Camurri, De Poli, Leman, & Volpe, 2005; Camurri et al., 2003) was adopted, including one low-level physical measure (velocity of the head movements) and a qualitative descriptor of gesture (quantity of motion).

Velocity of the head movements was calculated on the basis of the coordinates (*x, y*) of the barycenter (i.e., the centroid) of the pianist's head automatically extracted from the background.[3] Figure 1 (see color plate section) shows the automated tracking of the barycenter of the pianist's head. The module of velocity was computed taking into account its horizontal and vertical components (see Equation (1), where Δt is the time interval between subsequent video frames, 40 ms in this case).

$$|v| = \sqrt{v_x^2 + v_y^2}, \text{ where } \begin{cases} v_x(t) = \dfrac{x(t) - x(t-\Delta t)}{\Delta t} \\ v_y(t) = \dfrac{y(t) - y(t-\Delta t)}{\Delta t} \end{cases} \qquad (1)$$

Quantity of motion (QoM) is an approximation of the amount of detected movement, based on Silhouette Motion Images. A Silhouette Motion Image (SMI) carries information about all variations of the silhouette—a simplified view of the body consisting of a white color blob—shape and position in the last few frames (see Figure 2 in color plate section). The pianist's silhouette was obtained by means of computer vision techniques for background subtraction so that, in a black and white image, the pianist's silhouette is white and the background is black (see Figure 2).

$$SMI[t] = \left\{ \sum_{i=0}^{n} Silhouette[t-i] \right\} - Silhouette[t] \qquad (2)$$

The SMI at frame *t* is generated by adding together the silhouettes extracted in the previous *n* frames and then subtracting the silhouette at frame *t* (see Equation (2))—in such a way that only motion is considered, while the current posture is skipped. Thus, SMI carries information about the amount of motion that has occurred in the last *n* frames. QoM is computed as the area (i.e., number of pixels) of a SMI, normalized in order to obtain a value usually ranging from 0 to 1 (see Equation (3)). It can be considered as an overall measure of the amount of detected motion, involving velocity and force.

$$QoM = \frac{Area(SMI[t,n])}{Area(Silhouette[t])} \qquad (3)$$

Automated extraction allows for a temporal series to be obtained for each of the selected motion cues (QoM and velocity of the head movements) over time, depending on the video frame rate. Extraction of

[3] For the sake of brevity, within the text, the velocity of the head movements will be frequently labeled simply as velocity.

motion cues from the pianist's movement was done in EyesWeb XMI (Camurri, Coletta, Varni, & Ghisio, 2007), and particularly by using the EyesWeb Expressive Gesture Processing Library (Camurri, Mazzarino, & Volpe, 2004). For each gesture identified through the segmentation procedure, a subset of features related to the two motion cues was extracted. Those features described the dynamics of the motion cues over time.

Given $\vec{y} = [y_1, \ldots, y_N]$ the vector with the sequence of values of QoM or velocity over time in a gesture (N = number of samples for QoM or velocity in a gesture) and $\vec{m} = [m_1, \ldots, m_M]$ the vector with the sequence of the maximum values of QoM or velocity in a gesture (M = number of relative maxima in a gesture) ordered over time, a number of features aiming to explain the temporal dynamics of QoM and velocity were defined. These features were computed within each gesture and can be grouped in four classes according to the aspects of the motion cues' temporal profile they describe. Measures describing characteristics of the temporal profile of motion cues can often be directly associated with specific characteristics of the underlying gesture. For example, a temporal profile of velocity or QoM characterized by high peaks of short duration is usually observed in an impulsive gesture. However, since mappings to gesture characteristics are not obvious nor sufficiently empirically tested, the discussion in the following is limited to the analysis of characteristics of motion cues.

The first class provides information about the slopes of a motion cue's temporal profile—for QoM—and its main peak—for Velocity—in their starting (attack) and ending (release) phases. This choice is due to the typical shape of the velocity profiles (see Figure 3 in color plate section). Velocity has no peaks until the central part of the shape of the gesture; on the contrary, the QoM has several peaks, for example due to the preparation of the gesture.

Initial slope. For QoM, this feature is computed as the slope of the straight line joining the first maximum value and its initial value (see Figure 4 in color plate section),

$$\begin{cases} \dfrac{(m_1 - y_1)}{\Delta t} & \text{if } m_1 \neq y_1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where Δt is the temporal distance between the two points. This feature refers to a measure of the impulsiveness of the attack of the motion cue's temporal profile.

Final slope. For QoM, this feature is computed as the slope of the straight line joining the last maximum value and its final value (Figure 4),

$$\begin{cases} \dfrac{(y_N - m_M)}{\Delta t} & \text{if } m_M \neq y_N \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where Δt is the temporal distance between the two points. This feature refers to a measure of the impulsiveness of the release of the motion cue's temporal profile.

Initial slope of the main peak. For velocity, this feature is related to the main peak of its temporal profile—the one containing the motion cue's absolute maximum. It is computed as the first derivative between two significant points of the peak—the absolute maximum of the motion cue within the gesture and the minimum value preceding it (Figure 3),

$$\frac{(M - \min_{pre})}{\Delta t} \tag{6}$$

where, $M = \max\{y_i, i = 1 \ldots N\}$, $\min_{pre}$ is the minimum value preceding the absolute maximum and Δt is the temporal distance between the two points. This feature refers to a measure of the impulsiveness of the attack of the motion cue's main peak.

Final slope of the main peak. For velocity, this feature is related to the main peak of its temporal profile. It is computed as the first derivative between two significant points of the peak—the absolute maximum of the motion cue within the gesture and the minimum value following it (Figure 3),

$$\frac{(\min_{post} - M)}{\Delta t} \tag{7}$$

where $M = \max\{y_i, i = 1 \ldots N\}$, $\min_{post}$ is the minimum value following the absolute maximum and Δt is the temporal distance between the two points. This feature refers to a measure of the impulsiveness of the release of the motion cue's main peak.

The second class includes measures that describe the characteristics of the main peak of a motion cue's temporal profile, in terms of absolute value, temporal relevance, and duration. The main peak appears worthy of deeper consideration because it identifies the largest movement within a gesture, and, even more, when there is only one peak it practically represents the entire gesture. All the following features were calculated for both QoM and velocity.

Maximum value. This feature is related to the main peak of a motion cue and it is represented by the absolute maximum.

$$M = \max\{y_i, i = 1 \ldots N\} \tag{8}$$

Maximum value/Peak duration. This feature is related to the ratio between the maximum value of a motion cue within a gesture and duration of the main peak,

$$\frac{M}{PD} \tag{9}$$

where *PD* is the temporal duration of the main peak. This feature refers to the overall impulsiveness of a motion cue. A motion cue's temporal profile is impulsive when it is characterized by short peak duration with a high absolute maximum, while it is sustained when it is characterized by longer peak duration with a low absolute maximum.

Maximum value/Following maximum value. This feature is related to the relationship between the absolute maximum value of a motion cue within a gesture and the value of the biggest relative maximum (excluding the absolute maximum),

$$\frac{M}{M_1} \tag{10}$$

where $M_1 = \max\{y_i, i = 1 \ldots N, i \neq \arg M\}$. This feature explains how much the main peak is relevant with respect to the second one.

Main peak duration/Gesture duration. This feature is related to the relationship between the duration of the main peak of a motion cue and the duration of the gesture (see Figure 4),

$$\frac{PD}{GD} \tag{11}$$

where *PD* is the temporal duration of the main peak and *GD* the temporal duration of the gesture.

The third class includes three measures describing overall properties of a motion cue's temporal profile. All the following features were calculated for both QoM and velocity.

Mean value. This feature is related to the mean behavior of a motion cue and it is represented by the mean value over time of a motion cue within a gesture.

$$\mu = \sum_{i=1}^{N} \frac{y_i}{N} \tag{12}$$

Mean value/Maximum value. This feature is related to the relationship between mean and maximum value of a motion cue within a gesture.

$$\frac{\mu}{M} \tag{13}$$

Number of peaks. This feature refers to the number of peaks identified in a motion cue's temporal profile.

The last class includes features that describe the temporal regularity of the structure of a motion cue's profile. Temporal regularity refers to how regular and symmetric a motion cue's profile is in its development over time. All the following features were calculated for both QoM and velocity.

Centroid of energy. This feature gives an estimation of the barycenter of energy. The centroid of energy is the time instant corresponding to the time position of the barycenter of energy. For example, in the case of a motion cue with a strong attack and a smoother continuation, the centroid of energy will move toward the beginning of the motion cue's temporal profile. On the contrary, if energy is more concentrated in the end of a motion cue's temporal profile—a smooth attack with a strong conclusion—the centroid of energy will move toward the end of the motion cue's temporal profile. The centroid of energy is computed according to the following Equation:

$$\frac{\sum_{t=1}^{N} t y_t}{\sum_{t=1}^{N} t y_t} \tag{14}$$

where *t* is a time instant within the analyzed motion cue (*t* = 1 corresponding to the beginning of the motion cue's temporal profile, *t* = N corresponding to its end), and y_t is the value of a motion cue at the time instant *t*.

Distance between absolute maximum and the centroid of energy. This feature refers to the temporal distance between the absolute maximum of a motion cue within a gesture (*M*) and the centroid of energy. It also gives an indication on the relative position of the two points—the absolute maximum follows or precedes the centroid of energy in a motion cue's temporal profile.

Symmetry index. This feature gives an estimation of the symmetry of a motion cue's temporal profile. It is computed by means of the following steps: (1) identifying the center of the curve ($\bar{x}$), (2) calculating the difference between the areas below each half of the curve identified by the center, and (3) dividing by the total area below the whole curve to normalize between 0 and 1,

$$\frac{|\Delta t \sum_{i=\bar{x}}^{N} y_i - \Delta t \sum_{i=1}^{\bar{x}} y_i|}{\Delta t \sum_{i=1}^{N} y_i} = \frac{|\sum_{i=\bar{x}}^{N} y_i - \sum_{i=1}^{\bar{x}} y_i|}{\sum_{i=1}^{N} y_i}, \tag{15}$$

where y_i are the values of the motion cue (samples), Δt the temporal distance between the samples, ($\Delta t \Sigma_{i=\bar{x}}^{N} y_i$) the area below the right half of the curve with respect to the center, ($\Delta t \Sigma_{i=1}^{\bar{x}} y_i$) the area below the left half of the curve with respect to the center and the ($\Delta t \Sigma_{i=1}^{N} y_i$) total area below the whole curve.[4]

Shift index of the main peak. This feature gives an estimation of the position of the main peak of a motion cue with respect to its center. It is computed by means of the following steps: (1) calculating the position of the absolute maximum of the motion cue in the gesture, (2) subtracting the duration of the motion cue on the left from that one on the right with respect to the absolute maximum, and (3) normalizing with respect to the total duration of the gesture,

$$\frac{D_{right} - D_{left}}{GD}, \tag{16}$$

where GD is the duration of the gesture, D_{right} the duration of the motion cue's temporal profile on the right with respect to the position of the absolute maximum, and D_{left} the duration of the motion cue's temporal profile on the left with respect to the position of the absolute maximum.[5]

Number of peaks preceding the main one. This feature refers to the number of peaks preceding the one containing the absolute maximum.

Results

In order to examine the effects of the different expressive conditions on the player's performance, an analysis of variance (ANOVA) with repeated measures was conducted for each of the dependent variables described in the method section of this paper, with the 15 gestures located in the musical score treated as the random variable (15 values × 5 conditions, i.e., 75 instances) and the mode of expression as the independent variable (five levels). Means and standard deviations are reported in Table 1 (QoM related features) and in Table 2 (velocity related features).[6] The ANOVAs identified a significant main effect of the different modes of expression only on a small subset of the measured features. Significant effects emerged for four features related to the QoM and eight related to velocity.[7] For these 12 features, pairwise comparisons were performed (Bonferroni correction) in order to assess the specific difference among the emotionally expressive conditions.

For QoM, the following features showed a significant effect of expressive condition: *Final slope* [$F(4, 56) = 3.35$, $p < .05$], *Maximum value* [$F(4, 56) = 5.74$, $p < .01$], *Maximum value/Peak duration* [$F(4, 56) = 4.35$, $p < .01$], and *Mean value* [$F(4, 56) = 4.66$, $p < .01$]. Concerning velocity, the following features showed a significant effect of expressive condition: *Initial slope of the main peak* [$F(4, 56) = 3.56$, $p < .05$], *Final slope of the main peak* [$F(4, 56) = 4.96$; $p < .01$], *Maximum value* [$F(4, 56) = 7.63$, $p < .01$], *Mean value* [$F(4, 56) = 7.92$, $p < .01$], *Mean value/Maximum value* [$F(4, 56) = 5.53$, $p < .01$], *Main peak duration/Gesture duration* [$F(4, 56) = 4.10$, $p < .01$], *Maximum value/Peak duration*

[4] Area corresponds to the sum of the rectangles with Δt as basis and amplitude (y_i) as height. Symmetry index ranges from 0 (maximum symmetry) to 1 (minimum symmetry).

[5] If Shift Index is < 0, the main peak is overbalanced on the right of the curve; if Shift Index is > 0, the main peak is overbalanced on the left. If |Shift Index| = 0, the main peak is not overbalanced; if |Shift Index| = 1, the main peak is overbalanced.

[6] *Number of Peaks* and *Number of peaks preceding the main one*, for both motion cues, have been normalized and made continuous.

[7] Given the many ANOVAs and the limited sample, it was decided to adopt a partially conservative approach. Within the framework of an exploratory approach, it was decided to keep the significance level at $p = .05$ while being conservative in terms of ANOVA requirements, effect size, and power estimation. In the case of nonsphericity, effects were Greenhouse-Geisser corrected (the original degrees of freedom are reported throughout the article for readability reasons). Those features that showed a significant effect of the expressive mode were evaluated in terms of effect size and observed power. Three features (QoM *Centroid of energy*, QoM *Number of peaks preceding the main one*, and velocity *Symmetry*) were excluded from the discussion and pairwise comparisons because in all cases the observed power was extremely low (< .65).

TABLE 1. Mean Values and Standard Deviations of Quantity of Motion Related Features in the Different Expressive Conditions (N = 15 in each Condition; N Total = 75)

	Attack & Release		Main Peak				Overall			Gesture Regularity				
Quantity of of Motion	IS	FS	Max	Max/ PD	Max/ FMax	PD/ GD	Mean	Mean/ Max	NP	Centr	D Max Centr	Symm	Shift	NPP
Sad														
Mean	0.01	–0.01	0.13	0.01	1.52	0.24	0.05	0.37	0.29	25.92	8.54	0.30	–0.38	0.55
S.D.	0.01	0.01	0.06	0.01	0.62	0.11	0.02	0.12	1.33	12.66	10.79	0.24	0.43	1.40
Allegro														
Mean	0.02	–0.01	0.15	0.02	1.58	0.24	0.06	0.38	–0.05	20.94	1.52	0.31	–0.09	–0.08
S.D.	0.02	0.01	0.05	0.01	0.70	0.11	0.02	0.11	0.93	10.25	11.31	0.23	0.55	0.89
Serene														
Mean	0.02	–0.02	0.18	0.02	1.89	0.26	0.06	0.35	–0.21	19.39	–0.26	0.33	–0.04	–0.29
S.D.	0.01	0.02	0.06	0.01	1.14	0.13	0.03	0.06	1.07	10.85	11.45	0.20	0.55	0.78
Personal														
Mean	0.03	–0.02	0.16	0.02	1.45	0.20	0.06	0.37	–0.05	21.22	1.51	0.32	–0.02	–0.06
S.D.	0.03	0.01	0.05	0.01	0.45	0.06	0.03	0.09	0.88	12.48	9.06	0.18	0.58	0.92
Overexpressive														
Mean	0.02	–0.02	0.15	0.02	1.68	0.20	0.05	0.36	0.02	20.35	–0.09	0.35	–0.14	–0.13
S.D.	0.02	0.02	0.05	0.01	0.77	0.10	0.02	0.14	0.76	8.84	13.69	0.26	0.72	0.82

Note. IS = Initial slope; *FS* = Final slope; *Max* = Maximum value; *Max/PD* = Maximum value/Peak duration; *Max/FMax* = Maximum value/Following maximum value; *PD/GD* = Peak duration/Gesture duration; *Mean* = Mean value; *Mean/Max* = Mean value/Maximum value; *NP* = Number of peaks; *Centr* = Centroid of energy; *D Max Centr* = Distance between maximum and centroid of energy; *Symm* = Symmetry index; *Shift* = Shift index of the main peak; *NPP* = Number of peaks preceding the main one.

TABLE 2. Mean Values and Standard Deviations of Velocity of the Head Movements Related Features in the Different Expressive Conditions (N = 15 in each Condition; N total = 75)

	Attack & Release		Main Peak				Overall			Gesture Regularity				
Velocity of the head movements	ISP	FSP	Max	Max/ PD	Max/ FMax	PD/ GD	Mean	Mean/ Max	NP	Centr	D Max Centr	Symm	Shift	NPP
Sad														
Mean	39.95	−34.68	124.01	34.94	1.59	0.09	45.14	0.39	0.42	26.09	2.77	0.15	−0.06	0.36
S.D.	18.73	18.59	62.65	14.67	0.86	0.08	23.77	0.12	1.13	13.02	13.04	0.13	0.55	1.25
Allegro														
Mean	54.62	−59.29	195.31	37.30	1.72	0.12	51.52	0.34	0.12	23.82	0.05	0.24	0.01	0.01
S.D.	34.60	50.11	165.09	16.30	1.06	0.07	21.95	0.12	1.03	13.14	10.22	0.12	0.52	0.98
Serene														
Mean	85.34	−103.00	338.91	54.16	2.16	0.18	81.68	0.28	−0.48	20.37	0.89	0.23	−0.01	−0.28
S.D.	52.32	62.42	178.70	25.36	1.12	0.11	37.18	0.11	0.83	12.31	7.99	0.17	0.44	0.73
Personal														
Mean	67.41	−93.67	220.89	49.67	2.08	0.12	55.82	0.28	0.02	20.76	−0.16	0.19	0.04	−0.12
S.D.	29.10	60.68	107.32	20.53	0.84	0.06	20.00	0.07	1.08	10.82	8.14	0.09	0.38	0.83
Overexpressive														
Mean	56.22	−68.63	179.52	41.38	1.62	0.10	45.47	0.34	−0.04	21.65	2.15	0.28	−0.02	0.03
S.D	47.04	71.18	149.63	26.73	0.74	0.03	14.94	0.13	0.80	12.53	13.51	0.19	0.62	1.14

Note. ISP = Initial slope of the main peak; *FSP* = Final slope of the main peak; *Max* = Maximum value; *Max/PD* = Maximum value/Peak duration; *Max/FMax* = Maximum value/Following maximum value; *PD/GD* = Peak duration/Gesture duration; *Mean* = Mean value; *Mean/Max* = Mean value/Maximum value; *NP* = Number of peaks; *Centr* = Centroid of energy; *D Max Centr* = Distance between maximum and centroid of energy; *Symm* = Symmetry index; *Shift* = Shift index of the main peak; *NPP* = Number of peaks preceding the main one.

[$F(4, 56) = 4.35, p < .05$], and *Number of peaks* [$F(4, 56) = 6.37, p < .01$].

These results as a whole indicated that both motion cues, by means of a number of features, were sensitive to emotional expression in music, even though QoM significant features were only four out of fourteen. In previous studies, QoM proved to be sensitive to expressive intentions of dancers (Camurri et al., 2003). Most likely, the movements performed by the pianist while playing were not large enough to reveal differences among conditions in most of QoM related features. On the other hand, velocity showed a significant effect of the different conditions on many of its related features, supporting previous findings and suggestions (Dahl & Friberg, 2007; Davidson, 1994). Given the number of features showing a main effect of the expressive condition, results have been grouped for discussion in the four classes defined earlier.

Concerning the initial (attack) and final (release) phases of the motion cues' temporal profile and their main peak, three out of four features showed a significant effect of emotionally expressive condition—*Final slope* for QoM, *Initial slope of the main peak*, and *Final slope of the main peak* for velocity. The importance of attack and release in discriminating performances has already been shown in a number of studies, but mainly in musical terms (as reported in Gabrielsson & Juslin, 2003). It can be hypothesized that the same occurs for the movements of the pianist, when the starting phase of the gesture has fewer musical constraints, and thus the dynamics of the movements can be more influenced by expressive intentions. The second class of measures referred to the main peak of the motion cues' temporal profile. Significant effect of emotionally expressive condition has been found on five out of eight features, namely *Maximum value*, *Maximum Value/Peak duration*—for both QoM and velocity—and *Main peak duration/Gesture duration*—only for velocity. This high number of significant features highlighted that the main peak of the motion cues' temporal profile was highly sensitive to expressive intentions. The main peak changed its shape across the expressive conditions, especially in terms of absolute value and steepness (i.e., impulsiveness; *Maximum value/Peak duration*). The main peak temporal relevance within each gesture (i.e., proportion of peak duration with respect to the whole gesture) has been found only for velocity. This result clearly reflects the different typical shape of the velocity profile with respect to the QoM profile, the first usually characterized by a single large peak (see Figure 3 and Figure 4).

The third class of measures included six general descriptors of the motion cues' temporal profile. A clear difference between QoM and velocity was apparent from these features. All the three features—*Mean value*, *Number of peaks*, and *Mean value/Maximum value*—showed a significant effect of emotionally expressive conditions for velocity while, for QoM, only *Mean value* showed the same effect. Gestures differed in the average values of the two motion cues, implying that they both provide emotionally relevant information about the movement of a pianist. The number of peaks was found sensitive to emotionally expressive condition only in velocity profiles, showing that the peaks for velocity do not correspond to the peaks displayed in QoM profiles. It seems that velocity peaks did not necessarily co-occur with the movement of the upper body. As already mentioned, playing a piano implies some specific constraints for movement and these concern mainly the body of the pianist, while the head moves more freely. A possible further explanation of this difference is that QoM peaks may correspond with musical score elements. In this direction body movements of the pianist would be timely directed by the mere execution. Nevertheless, such a speculation requires further investigation.

The fourth class of measures included all the remaining features, which are sensitive to the temporal regularity of the cues' profile. None of these features showed a reliably significant effect of expressive condition. This lack of significant results can be interpreted as if the general profiles of the two motion cues were made by events whose timing did not change as a consequence of the expressive condition. In other words, the temporal regularity, for both motion cues, seems more responding to the score execution. Some body movements happen only when the musical score requires them.

Many QoM features did not show any significant effect of emotional expression. Further research on these features should assess their importance for music related emotional expressions. According to the present findings, they don't seem related to the emotionally expressive nuances of a piano performance.

Summing up, the first objective concerned the evaluation of QoM and velocity of the head movements as potentially meaningful motion cues for assessing emotional expression in music performance. Results did not show the role of QoM clearly, because the mode of expression had a significant effect only on four QoM features. On the other side the relevance of head movements (Dahl & Friberg, 2007) has been confirmed and reinforced by the automated approach applied. In particular, the release of QoM appeared sensitive to the emotionally expressive mode, as did the main peak for velocity, in terms of both absolute value and its temporal structure. These findings show parallels with musical aspects identified as being sensitive to emotion (Gabrielsson & Juslin, 2003) and strengthen the need for further studies focused on the multimodal expression of music performers, in terms of acoustics and body movements. Present findings are of interest considering the kind of material analyzed. Piano players are generally very constrained in their execution in most parts of the body, although head and back movements usually remain more flexible. This is reflected by the current general findings, where movements under different conditions showed variations more in terms of the velocity of head movements than in QoM of the whole upper body.

The second objective of the present research was the identification of more specific differences between expressive conditions. Pairwise comparisons (Bonferroni correction) of different emotionally expressive conditions were performed with respect to the significant features previously identified. Pairwise comparisons showed many significant differences involving all the conditions in the velocity features, while QoM features distinguished only among sad, serene, and personally felt affect conditions.

In terms of QoM, the sad expressive condition was lower than the serene expressive condition in *Maximum value* (*MD* = –.05; *p* < .05) and *Mean value* (*MD* = –.02; *p* < .05). In line with this outcome, sad expressive conditions were also lower than personal expressive conditions in terms of *Maximum/Peak duration* (*MD* = –.01; *p* < .05) and *Mean value* (*MD* = –.01; *p* < .05). These results indicate that the sad expressive condition was characterized by a generally low level of body movement with respect to the serene expressive condition. The fact that the same significant difference has been found between sad and personally felt affect expressions showed that it was the sad condition which contained the difference. The sad expressive condition appeared to be a performance in which the general amount of movement was reduced. This finding is in line with results from traditional research on body movement and emotions, where sadness has been found with little movement activity and a small spatial extension of gestures (Wallbott, 1998). It should be noted that sad was the only negatively valenced condition included and at the same time the less active. Further research should clarify this ambiguity, by investigating if QoM is more related to activity per se or is influenced by valence. Unfortunately, no other significant pairwise comparisons were found for QoM features, limiting the generalization of results and the conclusions that can be drawn about this motion cue.

In terms of velocity, sad expressive condition had a higher value in *Final slope of the main peak* than serene (*MD* = 68.33; *p* < .05) and personally felt affect expressions (*MD* = 58.99; *p* < .05). This feature referred to the speed of the velocity drop at the end of the main peak. High values in the *Final slope of the main peak* means that the speed of the velocity drop is low, since the final slope of the main peak is, accordingly with its mathematical definition, always negative. The combination of this finding with the significant higher value in sad expressive conditions for *Mean / Maximum* than serene (*MD* = .11; *p* < .05) and personally felt affect expressive conditions (*MD* = .11; *p* < .01) showed that the longer duration of the peak ending slope is coherent with the value of the maximum (which is lower). It seems that in the sad condition, it was not just the velocity that was lower, but also the negative acceleration. In other words, movements in the sad expressive condition became progressively slower over a longer period of time than in serene and personally felt affect. Unfortunately, no specific measures of acceleration have been included in the present study, so this interpretation cannot be fully empirically supported.

All the other significant pairwise comparisons for velocity involved serene expressive condition. Namely, serene expressive conditions showed a significant higher velocity *Mean value* than sad (*MD* = 36.54; *p* < .01), allegro (*MD* = 30.63; *p* < .05), and overexpressive (*MD* = 36.22; *p* < .05). These differences indicate that velocity of the head movements was generally higher in serene expressive conditions than in all the other conditions—with the exception of personally felt affect. This finding was replicated in velocity *Maximum value*: serene expressive conditions showed a significant higher average maximum value than sad (*MD* = 214.90; *p* < .01), allegro (*MD* = 143.60; *p* < .05) and overexpressive (*MD* = 159.39; *p* < .05).

Differences between serene and sad expressive conditions were expected in terms of velocity. Wallbott (1998) found a low level of movement dynamics, energy, and power for sad expressive conditions. Even though present findings replicate the same outcome, in the actual sample it was not the sad condition that showed significantly slower head movements than the others, but it was the serene expression condition which showed the fastest head movements. These findings were quite unexpected, since overexpressive and allegro were considered more active conditions than serene, and mean velocity was thought to be related to the general activation level. It is possible that the pianist's interpretation shifted from a medium activation indicated for serenity to a more activated positive emotion such as joy. Unfortunately, no specific self-report measures are available to disentangle this interpretation, so it is not possible to draw any specific conclusion about this hypothesis.

Expressions in the serene condition were also significantly different from sad expressive conditions ($MD = -8.60$; $p < .01$) and allegro expressive conditions ($MD = -5.67$; $p < .05$) in terms of the *Number of peaks* for velocity. There were less velocity peaks in serene expressive conditions. This finding indicated that serene expressive conditions had generally more uniform movements, in terms of velocity. A different outcome concerning the number of peaks between QoM and velocity is in line with the idea that velocity peaks do not correspond to the QoM peaks, because the latter would mainly represent movements related to the score execution—thus, showing a limited degree of variability among expressive conditions.

Summing up the present findings, two expressive conditions were more clearly differentiated than the others. Firstly, sad expressive condition was characterized by a low level of movement. In this expressive condition the pianist moved significantly less with respect to when she was in the serene and personal conditions. Unfortunately, no significant difference emerged for all the other conditions, so it is not possible to clarify the expressive meaning of QoM features. Secondly, serene expressive condition was significantly the highest in terms of velocity. This big difference with respect to allegro and overexpressive conditions was quite unexpected. In particular, overexpressive was defined as a condition where the movements were exaggerated without any specific emotional characterization. This definition is not reflected in the present results, raising some concerns about the appropriateness of the taxonomy that has been used for the study.

Nevertheless, some relevant indications do emerge from the current findings. Firstly, the two motion cues were able to characterize the different expressions in terms of general properties. Considering the mean values of the two motion cues, it appears that serene expressive condition had high QoM and fast head movements, personal expressive condition had high QoM with medium fast head movements, sad expressive condition had low QoM and slow head movements, allegro and overexpressive expressive conditions had medium QoM and slow head movements. Nevertheless, these differences have only a descriptive value, and they should be considered just as general trends, mainly showing the potential relevance of the two motion cues in describing different emotionally expressive performances. Secondly, two features dealing with the temporal aspect of the motion cues—QoM *Maximum value/Peak duration* and velocity *Final slope of the main peak*—were found specifically relevant for distinguishing the sad expressive condition from others. Given that the sad condition was the only one defined as having a negative valence, a possible interpretation of these features is that emotional valence may be more easily reflected in the timing of the expression than in absolute or averaged measures. While the specific correspondence between valence and timing is a speculation that cannot be proven only on the basis of these results, the importance of timing in emotional expression in music is increasingly becoming a central topic for research. For example, Schubert (2002) investigated the relationship between musical features and perceived emotion by using a continuous response methodology and time-series analysis; Luck and Toiviainen (2006) examined the features of conductors' gestures with which ensemble musicians synchronize their performance over time; Vines and colleagues (Vines, Nuzzo, & Levitin 2005), in a study considering temporal dynamics in music, analyzed data drawn from time-varying processes, such as continuous tension judgments, movement tracking, and performance tempo curves. Even though replications of current study with more performances and expressive conditions are necessary, present findings suggest moving towards the inclusion of timing as a key issue for disentangling emotional expressions in music performance.

The third objective of the present study was to assess the possibility of classifying different emotionally expressive performances on the basis of movement alone and to evaluate the significance of the features for emotion discrimination, as described next.

Since approaches based on decision trees have already proved successful in previous studies (Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2004), a decision tree was adopted. The algorithm J48—an implementation of the C4.5 decision tree (Quinlan, 1993)—was provided by the software Weka (Witten & Frank, 2005). Three different data sets were used as input for the algorithm: (1) QoM features, (2) velocity features, and (3) QoM and velocity features fused together.

All sets of data were normalized and discretized. Further, a wrapper approach to feature subset selection (which allows for the evaluation of the attribute sets by using a learning scheme) was used in order to reduce the number of inputs to the classifiers and to find the features that maximize the performance of the classifier. A best-first search method in the forward direction was used. This method searches the space of feature subsets by greedy hill-climbing augmented with a backtracking facility; forward direction means that the search starts with the empty set of features and moves forward (Witten & Frank, 2005). Cross-validation was used to build and evaluate the model from data.

Automated classification of the five emotional expressions based on the different subsets of features did not prove successful. The three expressions which were more differentiated according to the definitions of their labels and the results of the analysis of variance (i.e., sad, serene, and overexpressive) were then isolated and tested in a subsequent automated classification. Results highlighted the major role played by velocity with respect to QoM and that the two motion cues (QoM and velocity), fused together, allow for the highest recognition rate to be obtained from the classifier (57.8 %).

The decision trees built during the training and test phases showed that the most significant features in the classification process are the following: *Initial slope* for QoM and *Mean value* and *Initial slope of the main peak* for velocity. *Initial slope* for QoM highly discriminated between sad and overexpressive conditions—overexpressive performance showed higher values of this feature than the sad performance, meaning that the pianist's QoM in the overexpressive condition were characterized by a more impulsive attack. *Mean value* for velocity distinguished between serene and overexpressive conditions: serene performance was, on average, faster than the overexpressive one. Finally, *Initial slope of the main peak* for velocity discriminated between sad and serene conditions, with serene condition showing higher values. These results seem to suggest that the attack of the motion cues' temporal profile and their main peak plays a relevant role in discriminating emotionally expressive conditions. Further, results show that movements in the serene condition were faster than in the overexpressive one, a result that is in line with the output of the pairwise comparisons.

While the interpretation of the decision trees highlighted a subset of significant features that were partially coherent—*Initial Slope* was not significant for the ANOVA—with what emerged also from the statistical analysis, the automated classification did not prove as robust. This may be due to the algorithm used, or simply to the fact that the training of a model needs large datasets, which were not available in this study due to the limited number of repetitions of the emotionally expressive performances by the pianist. Nevertheless, these results, in line with what has emerged from previous analyses, strengthen the suggestion that the timing of gestures plays a major role in emotional expression in music performances.

Conclusion

This paper investigated the dynamic variations of gestures used by a pianist to communicate emotional expressive intentions while playing (Dahl & Friberg, 2007; Di Carlo & Guaitella, 2004; Thompson et al., 2005). Drawing from the theoretical constructs outlined in the Brunswikian model (Brunswik, 1956; Juslin, 1995, 2000), the present research focused on the distal cues of emotion communication.

The general aim was to assess the feasibility of the newly developed approach and thus its suitability for examining whether the movements of a music performer changed in terms of motion qualities according to different emotional intentions. The current study is one of the first attempts to use automated video analysis to extract expressive descriptors of movement in music performance. Automated video analysis techniques were used to gain more reliable measures of movements, and continuous variables were considered. These measures were preferred because movements in music performance cannot be clearly segmented into meaningful units as it is for gestures accompanying speech (Scherer & Wallbott, 1985; Wallbott, 1998). Furthermore, it complements the proposed approach—based on the dynamics of gestures—allowing the inclusion of time in the description of emotional expressions. Temporal profiles of specific motion cues, namely quantity of motion of the body and velocity of the head movements, were analyzed. Starting from these cues, a set of features was derived.

Results showed that the two motion cues were sensitive to emotional expression, but not in a uniform fashion. In particular quantity of motion did not appear strongly influenced by the emotionally expressive condition, with the exception of sad expressions. Most likely, pianist movements are quite constrained by the particular characteristic of the instruments, so variations in quantity of motion are smaller than for dancers (Camurri et al., 2003). This has been probably reflected in the fact that only the least active emotion has been differentiated by QoM. However, some features revealed variations among gestures confirming that they can convey information about emotional expressions. Especially, the timing of motion cues, namely the attack and release of their temporal profile and main peak, appears to merit more extensive investigation, since a dynamic account of emotional expression in music performance is encouraged by present findings.

Further, the main peak of the motion cues' temporal profile resulted quite sensible to the emotional expression, in terms of absolute value but also of temporal characteristics. Future studies can concentrate more on this aspect of the motion cues' temporal profile, which seems to convey much of the information, and on searching for relationships between the peaks and the musical score.

Differences between specific expressive conditions were found especially for the velocity of head movements. This evidence confirms previous findings (Dahl & Friberg, 2007; Davidson, 1994) and suggests to analyze music performances with overt reference to the specific instrument and its specific constraints. Thus, current findings indicate that the velocity of head movements can be sensitive to a positive emotional valence of the expressive condition—serene was the fastest condition—but its role for musicians utilizing other instruments than piano may be different. Further research should concentrate on this motion cue, investigating more expressive categories and applying similar sets of features to musicians who play different instruments. This can give insights on whether velocity of the head movements is a typical expressive cue only for pianist performers.

The main limitation of the present study is that analyses are based on one musician playing one music excerpt. Even the characteristics of the musical piece can have an influence on emotionally expressive rendering—for example the tempo marking. Further, the high number of ANOVAs is likely to make the identification of spurious results more probable, even if this concern caused a more conservative approach to the analysis and results discussion. These limitations preclude a generalization concerning the cues and features extracted as well as the general effects of emotional expression on musical performance. Additional studies with larger numbers of musicians and excerpts should be conducted in order to assess the validity of the present results.

The present research highlights the difficulty of identifying effects on movement when depending exclusively on expressive intentions. Future research should address emotionally expressive movements taking into account also the effects on the movements due to the structural features of the music itself, as well as the way the latter interact with movements due to emotion expression only. In order to clarify this issue, it may be useful to jointly analyze the several modalities of expression that musicians use while performing, especially their facial expressions (as shown by Thompson et al., 2008). The combined analysis of different modalities of expression is likely to provide a clearer description of how musicians express emotions in their performances. A multimodal account of emotional expression in music should be based on a Brunswikian perspective (Brunswik, 1956), as suggested by Juslin (1995, 2000) and Scherer (1978, 2003), respectively for musical expression and for emotional expression. The notion of vicarious functioning (Brunswik, 1956; Juslin, 2000) may help to explain the variability of cues across modalities in different expressive performances and their use by listeners.

Finally, this research area is in urgent need of a refinement of the emotion taxonomy to be used in research on musical expression. As the classic basic emotion approach is of limited utility for music and as the Wundtian dimensions of valence and arousal are far too general (Scherer, 2004), new approaches need to be developed. In this study we used ad hoc categories defined in interaction with the musician. However, this study shows that the results are not always easy to interpret, as some of the categories are conceptually underspecified. Consequently, research efforts should be directed at the development of a taxonomy that is meaningful for music performers and emotion researchers alike.

Author Note

Ginevra Castellano is now at the Department of Computer Science, School of Electronic Engineering and Computer Science, Queen Mary, University of London, UK.

The research work has been conducted in the framework of the EU-IST project HUMAINE (Human-Machine Interaction Network on Emotion), a network of excellence (NoE) in the EU 6th Framework Programme (2004-2007), and partially supported by the National Center of Competence in Research (NCCR) Affective sciences financed by the Swiss National Science Foundation (n° 51NF40-104897) and hosted by the University of Geneva. The authors also acknowledge the major contribution of Tanja Bänziger and Daniel Arfib in organizing and conducting the recording and reviewers for useful comments. In particular, we thank Esther Walker, the pianist, for her enthusiastic participation in this study. We also thank Christopher Peters for proofreading the English in the manuscript.

Correspondence concerning this article should be addressed to Ginevra Castellano, Department of Computer Science, School of Electronic Engineering and Computer Science, Queen Mary, University of London, Mile End Road, London E1 4NS, UK. E-MAIL: ginevra@dcs.qmul.ac.uk

References

Argyle, M. (1980). *Bodily communication*. London: Methuen & Co Ltd.

Balomenos, T., Raouzaiou, A., Ioannou, S., Drosopoulos, A., Karpouzis, K., & Kollias, S. (2005). Emotion analysis in man-machine interaction systems. In S. Bengio & H. Bourlard (Eds.), *Lecture notes in computer science: Vol. 3361. Machine learning for multimodal interaction* (pp. 318-328). Berlin: Springer-Verlag.

Bernhardt, D., & Robinson, P. (2007). Detecting affect from non-stylised body motions. In A. Paiva, R. Prada, & R. W. Picard (Eds.), *Lecture notes in computer science: Vol. 4738. Affective computing and intelligent interaction, second international conference* (pp. 59-70). Berlin: Springer Verlag.

Boone, R. T., & Cunningham, J. G. (1998). Children's decoding of emotion in expressive body movement: The development of cue attunement. *Developmental Psychology, 34*, 1007-1016.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Camurri, A., Coletta, P., Varni, G., & Ghisio, S. (2007). Developing multimodal interactive systems with EyesWeb XMI. *Proceedings of the 2007 conference on new interfaces for musical expression (NIME07)* (pp. 305-308). New York, USA.

Camurri A., De Poli G., Leman M., & Volpe G. (2005). Toward communicating expressiveness and affect in multimodal interactive systems for performing art and cultural applications. *IEEE Multimedia Magazine*, *12*, 43-53.

Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies*, *59*, 213-225.

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2004). Multimodal analysis of expressive gesture in music and dance performances. In A. Camurri & G. Volpe (Eds.), *Lecture notes in artificial intelligence: Vol. 2915. Gesture-based communication in human-computer interaction* (pp. 20-39). Berlin: Springer Verlag.

Camurri, A., Mazzarino, B., & Volpe, G. (2004). Analysis of expressive gesture: The Eyesweb expressive gesture processing library. In A. Camurri & G. Volpe (Eds.), *Lecture notes in artificial intelligence: Vol. 2915. Gesture-based communication in human-computer interaction* (pp. 460-467). Berlin: Springer Verlag.

Canazza S., De Poli G., Drioli C., Rodà A., & Vidolin A. (2000). Audio morphing different expressive intentions for multimedia systems. *IEEE Multimedia Magazine*, *7*, 79-83.

Castellano, G, Villalba, S. D., & Camurri, A. (2007). Recognising human emotions from body movement and gesture dynamics. In A. Paiva, R. Prada, & R.W. Picard (Eds.), *Lecture notes in computer science: Vol. 4738. Affective computing and intelligent interaction, second international conference* (pp. 71-82). Berlin: Springer Verlag.

Clarke E. (1988). Generative principles in music performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition*, (pp. 1-26). New York: Oxford University Press.

Cooke, D. (1959). *The language of music.* London: Oxford University Press.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musician's body movements. *Music Perception*, *24*, 433-454.

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.

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 Poli G., Rodà A., & Vidolin A. (1998). Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance. *Journal of New Music Research*, *27*, 293-321.

Di Carlo N. S., & Guaitella, I. (2004). Facial expressions of emotion in speech and singing. *Semiotica*, *149*, 37-55.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, *29*, 288-298.

Friberg, A. (1995). *A quantitative rule system for musical performance*. Unpublished doctoral dissertation, Royal Institute of Technology, Stockholm, Sweden.

Gabrielsson, A., & 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.

Gabrielsson A., & Juslin P. (2003). Emotional expression in music. In R. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 503-534). New York: Oxford University Press.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 223-248). New York: Oxford University Press.

Gunes, H., & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, *30*, 1334-1345.

Juslin, P. N. (1995). Emotional communication in music viewed through a Brunswikian lens. In G. Kleinen (Ed.), *Musical expression. Proceedings of the conference of ESCOM and DGM 1995* (pp. 21-25). Bremen, Germany: University of Bremen.

JUSLIN, P. N. (1997a). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, *14*, 383-418.

JUSLIN, P. N. (1997b). Can results from studies of perceived expression in musical performances be generalized across response formats? *Psychomusicology, 16*, 77-101.

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*, 1797-1813.

JUSLIN, P. N., & LAUKKA, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*, 217-238.

KAHOL, K., TRIPATHI, P., PANCHANATHAN, S., & RIKAKIS, T. (2003). Gesture segmentation in complex motion sequences. *Proceedings IEEE International Conference in Image Processing* (pp. 105-108), Barcelona, Spain.

LABAN, R. (1963). *Modern educational dance*. London: Macdonald & Evans Ltd.

LABAN, R., & LAWRENCE F. C. (1947). *Effort*. London: Macdonald & Evans Ltd.

LANGER, S. K. (1957). *Philosophy in a new key* (3rd ed.). Cambridge, MA: Harvard University Press.

LIVINGSTONE, S. R., & THOMPSON, W. F. (2006). Multimodal affective interaction. A comment on musical origin. *Music Perception*, *24*, 89-94.

LUCK, G., & TOIVIAINEN, P. (2006). Ensemble musicians' synchronization with conductors' gestures: An automated feature-extraction analysis. *Music Perception*, *24*, 189-199.

POLLICK, F. E., PATERSON, H., BRUDERLIN, A., & SANFORD, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*, B51-B61.

QUINLAN, R. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann Publishers.

SCHERER, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology*, *8*, 467-487.

SCHERER, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92-120). New York and Oxford: Oxford University Press.

SCHERER, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227-256.

SCHERER, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research*, *33*, 239-251.

SCHERER, K. R., & WALLBOTT, H. G. (1985). Analysis of nonverbal behavior. In T. A. Van Dijk (Ed.), *Handbook of discourse analysis* (pp. 199-230). London: Academic Press.

SCHERER K. R., & ZENTNER M. R. (2001). Emotional effects of music: Production rules. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 361-392). Oxford: Oxford University Press.

SCHUBERT, E. (2002). Correlation analysis of continuous emotional response to music: Correcting for the effects of serial correlation. *Musicae Scientiae*, *Special Issue 2001-2002*, 213-236.

SLOBODA, J. A., & JUSLIN, P. N. (2001). Psychological perspectives on music and emotion. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 71-104). New York: Oxford University Press.

SUNDBERG, J., FRIBERG, A., & FRYDÉN, L. (1991). Common secrets of musicians and listeners: An analysis-by-synthesis study of musical performance. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 161-197). London: Academic Press.

THOMPSON, W. F., SUNDBERG, J., FRIBERG, A., & FRYDÉN, L. (1989). The use of rules for expression in the performance of melodies. *Psychology of Music*, *17*, 63-82.

THOMPSON W. F., GRAHAM P., & RUSSO F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, *156*, 203-227.

THOMPSON, W. F., RUSSO, F. A., & QUINTO, L. (2008). Audio-visual integration of emotional cues in song. *Cognition and Emotion*. Retrieved May 27, 2008, from http://www.informaworld.com/smpp/ftinterface~content=a789794895~fulltext=713240930.

TIMMERS, R., MAROLT, M., CAMURRI, A., & VOLPE, G. (2006). Listeners' emotional engagement with performances of a skriabin etude: An explorative study. *Psychology of Music, 34*, 481-510.

VINES, B., WANDERLEY, M. M., NUZZO, R., LEVITIN, D., & KRUMHANSL, C. (2004). Performance gestures of musicians: What structural and emotional information do they convey? In A. Camurri & G. Volpe (Eds.), *Lecture notes in artificial intelligence: Vol. 2915. Gesture-based communication in human-computer interaction* (pp. 468-478). Berlin: Springer Verlag.

VINES, B., NUZZO, R. L., & LEVITIN, D.J. (2005). Analyzing temporal dynamics in music: Differential calculus, physics, and functional data analysis techniques. *Music Perception*, *23*, 137-152.

WALLBOTT, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, *28*, 879-896.

WALLBOTT, H. G., & SCHERER, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, *51*, 690-699.

WITTEN, I. H., & FRANK, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco: Morgan Kaufmann.