Towards the Automated Analysis of Expressive Gesture Qualities in Full-Body Movement: The Perceived Origin of Movement

Olga Matthiopoulou, Giorgio Gnecco, Marcello Sanguineti, Denis Mottet, Benoit G. Bardy, Antonio Camurri

Human-centric Computing and Information Sciences

September 2024 | Volume 14

KCIA
Korea Computer Industry Association

www.hcisjournai.com

HCIS Human-centric Computing and Information Sciences

RESEARCH (Research Manuscript) Open Access

Human-centric Computing and Information Sciences (2024) 14:54
DOI: https://doi.org/10.22967/HCIS.2024.14.054
Received: February 20, 2023; Accepted: July 16, 2023; Published: September 30, 2024

Towards the Automated Analysis of Expressive Gesture Qualities in Full-Body Movement: The Perceived Origin of Movement

Olga Matthiopoulou[1], Giorgio Gnecco[2,*], Marcello Sanguineti[1], Denis Mottet[3], Benoît Bardy[3], and Antonio Camurri[1]

Abstract

The automatic detection of the perceived origin of full-body human movement (OoM), i.e., of the part of the body that an external observer perceives as the joint where the movement originates, is a relevant topic for human movement analysis, as it can allow one to interpret affective content and social signals and can have applications in cognitive/motor rehabilitation, among others. Within this framework, the objective of this work is to present a computational method aimed at the automatic detection of the perceived OoM, starting from movement features acquired via motion capture techniques. After defining the concept of perceived OoM, the following contributions are presented: a set of techniques for the automated analysis of full-body expressive non-verbal communication, based on several low-level local movement features of the joints (speed, tangential acceleration, angular momentum); a computational method for the automatic detection of the perceived OoM at different spatial scales; and a repository of full-body movements annotated in terms of the perceived OoM, adopted for validation and evaluation of the method. The results of the analysis demonstrate its effectiveness. Finally, possible extensions of the method are outlined.

Keywords

Non-verbal Communication, Full-Body Movement Analysis, Automated Analysis of the Perceived Origin of Movement, Graph Theory, Cooperative Game Theory

1. Introduction

Humans make inferences on the affective meaning of full-body non-verbal communication. Postural attitudes and full-body movement are of paramount importance in conveying affective content [1–3]. Furthermore, the relationship between motion and emotion, i.e., how individual emotions propagate through embodied presence in a group, and how joint action changes individual emotions, represents another challenging research direction to understand and model affect [4–8]. Recent results in the analysis of expressivity and emotion from full-body movement, together with the availability of low-cost interactive multimodal movement technology, are leading the human-computer interaction community

***Corresponding Author:** Giorgio Gnecco (giorgio.gnecco@imtlucca.it)
[1]Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genova, Genova, Italy
[2]AXES Research Unit, IMT School for Advanced Studies Lucca, Lucca, Italy
[3]EuroMov Digital Health in Motion, University of Montpellier, Montpellier, France

to focus on embodiment and the role of full-body in interaction design [9]. In such a context, this paper presents a computational method for the automated analysis of expressive qualities in full-body human movement: specifically, it focuses on the detection of the perceived origin of movement (OoM).

The perceived OoM is the specific part of the body that an external observer perceives as the joint from which movement originates. The capacity of an observer to understand, moment by moment, the part of the body where movement originates is an important cue that can contribute to focusing her attention and interpreting affective content and social signals. For example, consider a direct forward movement of the arm from one person to another: a "caressing" gesture will be characterized by the hand as the (perceived) OoM; a "rejection" gesture will probably have the shoulder as the OoM, while an aggressive movement (e.g., a "punch") may have the right or left foot as the OoM, to convey energy from the ground to the hand through the body kinematic chain foot-hand. Moreover, the analysis of the OoM is an important component in guiding therapists for individualized patient care. For example, it has been shown in [10, 11] that diagnosing the origin of a reaching movement is very useful for adapting the rehabilitation of a person with a stroke. In cognitive/motor rehabilitation, the automated detection of the OoM can help a patient learn how to perform a specific movement correctly (e.g., how to get up safely from a chair), thus reducing the risk of injury. Finally, in dance, music, and sports, awareness and discovery of the OoM can contribute to enhanced performance and effectiveness of expressivity.

It follows from the discussion above that the development of methods to automatically detect the perceived OoM is relevant for a better understanding of human movement. In this context, the present article describes a computational method to automatically detect the perceived OoM, extending the previous work [12], coauthored by a subset of its authors. The method takes as inputs full-body movement features computed at the physical level following a motion capture (MoCap) data acquisition. After defining the concept of OoM, the article presents (1) a set of techniques for the automated analysis of low-level movement features of joints in expressive full-body non-verbal communication; (2) a computational method for the automatic detection of the perceived OoM, obtained by combining tools and techniques from graph theory and cooperative game theory, starting from low-level movement features; and (3) a repository of full-body movements, annotated in terms of the perceived OoM, created for the validation and evaluation of the method.

One first proposal for a computational method to detect the perceived OoM was made in [12]. To the authors' knowledge, the problem of the automatic detection of the perceived OoM was not investigated in any other recent journal article. The method of [12] is based on a mathematical game built over a graph structure representing the human body. The aim is to extract from this model information about the full-body movement performed by a subject. The work [12] is taken as a starting point to further develop this computational method for the analysis of expressive qualities of full-body movement.

The main contributions of this paper are as follows.

- In contrast to the previous work [12], the detection of the perceived OoM here is based on the use of a larger set of movement features that are computed for each body joint (by considering speed, tangential acceleration, and angular momentum). These features are the input data to the game-theoretic component of the method, which is solved in terms of the Shapley value. This is a well-known solution concept in cooperative game theory [13], which measures the importance of each player in a cooperative game of appropriate form.
- Here, the method includes an additional feature (the mass distribution feature), which is used to assess the quality of the approximation made by converting the original graph structure of the human body into a reduced one with a smaller number of vertices, which is more suitable for the construction of ground truth. This is also an important contribution towards future real-time implementation of the method.
- A comprehensive evaluation of the computational method is provided, together with its comparison with alternative methods. The discussion of the method includes a thorough analysis of the available ground truth, by relating the confidence values expressed by the participants in an online survey

about the evaluation of the perceived OoM to their inter-participant agreement, as one of the aims of the investigation made in the present work is to detect fragments with a clearly perceived OoM.

- Furthermore, as an improvement in the implementation of the computational method, the present work includes a more efficient pre-processing of the MoCap data, obtained by removing outliers, considering biomechanical constraints of human movement, and then filtering the data to further reduce the amount of noise.

The paper is structured as follows. Section 2 provides a background on the analysis of expressive human movement, and on the literature that motivates studying the concept of perceived OoM. Section 3 presents a definition of the perceived OoM and a computational method for its automatic detection. Section 4 describes the annotated dataset developed to evaluate and validate the computational method. Section 5 provides details on the MoCap data and its pre-processing. Section 6 illustrates the system architecture of the computational method presented in Section 3, focusing on the definition of the selected low-level movement features. Section 7 focuses on the validation of the method, and Section 8 reports the results of the analysis. Section 9 concludes the paper and discusses possible future research directions. A preliminary version of Sections 5, 6.2, and 7 was presented in the short conference paper [14].

2. Background

2.1 Bodily Expression of Affect

Research on computational methods for human movement benefits from a fruitful "scientific contamination," which integrates biomechanics and neuroscience, experimental psychology, as well as, in the case of expressive movements, theories from the arts and humanities [15, 16]. Several studies have investigated the role of full-body expression as a tool for communication. However, research on full-body expression is still scarce. In [17], the authors highlighted the importance of including bodily expressions in affective neuroscience. Its author argued that bodily expressions should be as thoroughly researched as facial expressions since both are important components for the recognition of emotions. Facial expressions are connected to one's mental state, whereas bodily expressions draw attention to individual or group actions. As a result, one person can infer the affective state of another not just by reading her face, but also by reading her actions. Moreover, according to [17], bodily expressions are just as well perceived in a multi-sensory way as facial expressions. In essence, the perception of an affective state is multimodal.

The authors of [18] examined the role of full-body expressions in communication, concluding that these types of full-body expressions are indeed able to carry meaning and information, to a greater extent than previously thought for non-verbal communication. The article described the two separate pathways in the brain—namely, form and motion information—that play a role in the recognition of biological information. This showed that the perception of affect from full-body expressions depends heavily on both form and motion information.

Furthermore, examining body movements for affective expression, [19] found that body movements can successfully convey affective expressions in multiple manners, such as through whole-body gestures, more isolated movements such as arm gestures, or modulation of functional movements.

The authors of [20] arrived at some interesting conclusions when considering how emotions are recognized through dance movements. The analysis in that article was based on five dancers performing the same dance but conveying four different emotions, namely anger, fear, grief, and joy. Its results showed that the intended emotions were clearly recognized through the dance movements, with the stronger negative emotions (grief and anger) being the most recognizable. It follows that the development of a computational feature-based analysis of body movements should be based on a close link between research in cognitive neuroscience and full-body movement analysis.

Finally, the authors of [21] found that the brain organizes and categorizes human movement in such a way as to extract meaning from it. For what concerns the perception of movement features, several brain

regions are connected to different functions. The article found that, rather than semantically structuring the perception of bodily motions, the brain tends to organize it according to shape and movement qualities. In particular, low-level features are related to activations in regions of the early visual hierarchy as well as in motion-sensitive regions. On the other hand, mid-level features are linked to postural attributes encoded in other parts of the brain.

2.2 The Leading Joint Hypothesis

Another source of inspiration for the approach used in this article comes from movement science and biomechanics, and in particular from the literature on the leading joint hypothesis (LJH) of limb motion: "There is one leading joint that creates a dynamic foundation for motion of the entire limb" [22]. The basis of the LJH is found in the way the central nervous system exploits the biomechanical properties of the limbs for movement organization. Since human limbs have a multi-joint structure, this causes interactions between segments that are motion-dependent. Specifically, due to the articulated structure of the limbs, each joint is responsible for different functions in the production of a movement. However, there is often one joint that leads the movement, due to its placement within this linkage structure. Consider a punch, where one can see both the arm and the corresponding hand in motion. While these two parts of the body are moving, this does not necessarily classify one of them as the OoM. Indeed, the energy and strength used to throw a punch can come from either a foot or a shoulder joint, thus classifying either one of these as the leading part of the body in this situation. The author of [22] also made a notable distinction between the role of proximal and distal joints in the LJH. Because of the greater mechanical influence of proximal joints on the motion, the leading joint is more frequently a proximal joint than a distal joint. Nevertheless, context and task are also among the major determinants of the leading joint. For example, a simple hand wave compared to a full-arm wave requires a much smaller range of motion at the proximal joints compared to the distal joints, resulting in an insignificant mechanical effect of the proximal joints. The former gesture requires a full range of motion at the wrist, whereas the latter requires a full range of motion at the shoulder joint.

3. Theoretical Framework

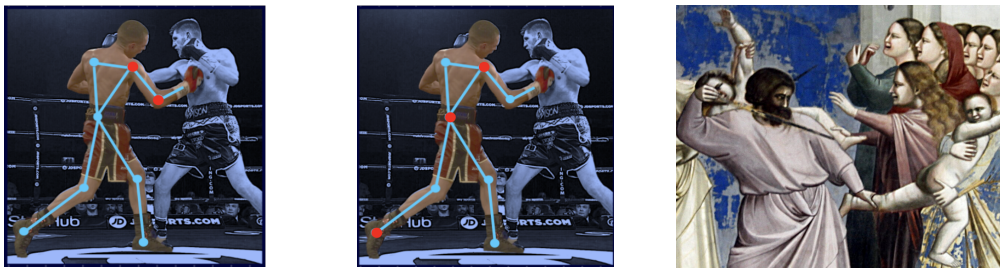
3.1 Origin of Movement – A Definition

Analyzing and understanding the OoM is a powerful way to model non-verbal affective behavior and social signals. In this article, the OoM is defined as the joint where the movement originates and begins to propagate through the body to other joints. The OoM characterizes the meaning of a movement, as in the previous example of the different ways an arm forward extension can be perceived: a punch originates from the foot, a rejection originates from the shoulder, while a gentle caress originates from the hand. The difference in the dynamics of these actions creates a distinction as to where a movement is initiated. It is possible to distinguish two perspectives in the definition of the OoM. First, the perceived OoM refers to the viewpoint of an external observer and is defined as the part of the body from which a movement appears to originate. Second, the acted OoM refers to the biomechanical level, i.e., to how the muscle works at the leading joint. The assumption made in this paper is that an observer is quite capable of identifying the true leading joint, i.e., that the perceived and acted OoM do not differ that much. Of course, an action can be performed to induce the perception of a fake OoM, but this interesting special case is not examined in this paper.

Consider the following example, shown in Fig. 1. Depending on how the motor plan is conceived, there are different ways to execute the gesture of a punch, each characterized by a different OoM. In an incorrect execution (i.e., ineffective, resulting in a low-force punch), the punch originates from the

shoulder (Fig. 1(a)): as a result, this execution has a low impact on the opponent. In a correctly executed punch, the movement originates from the foot, and the energy flows and grows from the foot through the kinematic chain of the whole body moving up through the hip to the shoulder and hand, resulting in a more effective (i.e., stronger, more aggressive) punch gesture (Fig. 1(b)). This illustrates the role of the OoM in explaining expressive intent and affective communication.

Another example of the correlations between OoM and emotion is the painting shown in Fig. 1(c). There, the mother's arm is moving upward: how can one distinguish in her gesture the tender sweetness leading her to caress the cheek of her sleeping child, or the inner violence that is preparing to hit the soldier's cheek? If one tries to imagine the movement of the arm, then the perception of the OoM in the hand may contribute to recognizing a caress, while the perception of the OoM in the shoulder may explain an aggressive hit to the soldier's cheek. Similarly, still referring to the same painting, how can one foresee in the heavy murderer's hand movement a slight vacillation revealing his fragile uncertainty in completing his task (therefore with a time-varying OoM between different joints)? How to measure the hesitation of the woman's arms closing to protect her crying baby, in resonant dialog with the bodies moving around her? In conclusion, measuring the OoM is expected to provide one of the most important cues to explain complex affective intentions and social signals.



(a) (b) (c)

Fig. 1. Examples of gestures and related perceived origins of movement (OoMs). (a, b) Two possible and different OoMs for a punch (images taken from https://boxingscience.co.uk/science-behind-punch). (c) Detail from Giotto's painting "La strage degli innocenti" (Italian title, ca. 1303-1305) in Cappella degli Scrovegni, Padua, Italy (see Section 3.1 for a discussion on the perceived OoM).

3.2 The Computational Method

This research is based on the previous work [12], in which an evaluation of human movement qualities was performed by using a method based on transferable-utility games on graphs, and in particular on a game-theoretic solution concept known as the Shapley value. In the present context, the Shapley value quantifies the importance of each joint in the generation of a human full-body movement. Combining graph and game theory provides a novel approach to movement analysis, in that it uses a game-theoretic component for the identification of the most important joint, and it exploits a graph-theoretic component that facilitates the representation of the skeletal structure of the body as a graph. In the following, the method developed in [12] is summarized. The extensions made in the present article are detailed in the next sections.

The core of the method of [12] is based on a suitable quantification of the concept of joint importance in a movement. Loosely speaking, that method searches for the most important joint (more specifically, the most important vertex in a suitable graph representation of the human body), as the one connecting two (or more than two) different automatically generated clusters. These clusters, as described later, are defined in terms of joints that are connected and have similar movement qualities. The steps of the method are summarized as follows (the reader is referred to [12] for a more detailed description).

Step 1. A weighted undirected graph $G = (V, E, w)$ is constructed to model the human body through its skeletal structure, where V is the set of vertices, E is the set of edges, and w is a weight function

defined on E. The vertices of such a graph form a subset of joints of the body, whereas its edges are either physical edges or non-physical edges. The former are actual edges of the skeletal structure. The latter, also called "bridge" edges, connect vertices that are not physically linked. More precisely, non-physical edges model the temporary similarity in the movement performed by parts of the body that are not directly connected by an edge within the skeletal structure.

In the method developed in [12], for each frame, the physical edges are assigned non-negative weights proportional to the current (non-negative) similarity of the values assumed by a given movement feature at each of the two vertices associated with such edges. In that work, speed was chosen as this feature for simplicity. Non-negative weights are also assigned to non-physical edges, again proportional to the current similarity in the feature values of the associated vertices. The constant of proportionality is chosen to be much smaller in the case of non-physical edges since these edges should only be associated with a large weight if there is a very large similarity of feature values between the two vertices.

Step 2. The next step of the method is based on the application of a clustering technique known in the literature as spectral clustering, which is suitable for clustering data represented by graphs [23, 24]. For each frame, the weighted undirected graph G defined above is clustered by applying spectral clustering to its weighted edge set. This allows for vertices of the same automatically detected cluster to have quite similar feature values, and for edges between different clusters to be associated with vertices that have quite dissimilar feature values.

Step 3. Subsequently, for each frame, an appropriate weighted auxiliary graph is defined, denoted as $G^{aux} = (V, E^{aux}, w^{aux})$. The vertices of G^{aux} are the same as in the original graph G. In contrast, the edge set of G^{aux} is a subset of physical edges of G, which connect vertices belonging to different clusters of G. The weight of each edge in G^{aux} is proportional to the dissimilarity of the values assumed by the selected movement feature on its two associated vertices (unlike the original graph G, for which the similarity is considered, not the dissimilarity). Note that in the auxiliary graph G^{aux}, non-physical edges are no longer considered, to give more importance to physical edges in this phase of the analysis (in this way, in the successive steps, one avoids estimating the OoM as being located between two vertices that are not connected by a physical edge, i.e., one gives more importance to the physical edges).

Step 4. As a further step, for each frame, a cooperative transferable utility (TU) game is constructed on the weighted auxiliary graph G^{aux}. TU games (see [13] for more details) are defined in terms of a set of players N and of a characteristic (or value) function, whose argument is a generic subset of N (also called coalition). In the method developed in [12], the players are the vertices of G (or equivalently, of G^{aux}). Thus, in this case, the set of players of the TU game coincides with the set of vertices of G (or of G^{aux}), i.e., $N = V$. The value $c(V')$ of a generic coalition $V' \subseteq V$ is chosen as the summation of all the weights (in the weighted auxiliary graph G^{aux}) associated with the physical edges belonging to the subgraph of G^{aux} that is induced by V'. In other words, the characteristic function of the TU game is set as $c(V') = \sum_{v,\hat{v} \in V', \hat{v} \in N^{aux}(v)} w^{aux}\left(e^{aux}_{v,\hat{v}}\right)$, where $N^{aux}(v)$ denotes the set of neighbors of a vertex v in the weighted auxiliary graph G^{aux}, and $w^{aux}\left(e^{aux}_{v,\hat{v}}\right)$ is the weight of the edge $e^{aux}_{v,\hat{v}}$ connecting the two vertices $v, \hat{v} \in V'$.

Step 5. Then, for each frame, the Shapley value for the cooperative TU game constructed in Step 4 is computed. The Shapley value is a popular solution concept in game theory. It is a measure of the value of each player in a TU game [13]. The general idea behind the Shapley value is that the importance of each player equals its average marginal contribution to the utility of a coalition when the player joins that coalition (the average being computed with respect to a suitable probability measure on the set of such possible coalitions). The Shapley value represents a fair way to allocate to the players the utility of the coalition formed by all of them (this is called the grand coalition) and satisfies a well-known axiomatic characterization, i.e., a set of desirable properties for such a fair allocation (see again [13] for details). More precisely, for each player $i \in V$, its Shapley value $\varphi_i(c)$ in the TU game with characteristic function c is defined as $\varphi_i(c) = \sum_{V' \subseteq V \setminus \{i\}} \frac{|V'|!\,(|V|-|V'|-1)!}{|V|!}\left(c(V' \cup \{i\}) - c(V')\right)$. Here, the term $(V' \cup \{i\}) - c(V')$ represents the marginal contribution of player i when it joins a coalition V' (which does not

include i), whereas $\frac{|V'|!\,(|V|-|V'|-1)!}{|V|!}$ is a weight attributed to this marginal contribution, which depends on the number of players $|V|$ and the size $|V'|$ of V'.

In the case of a TU game defined on a network in which the players are its vertices (or arcs), the Shapley value represents a means to evaluate the "centrality" or "importance" of such vertices [25, 26] (or arcs [27]). For some games, its exact evaluation is computationally demanding [28], but this is not the case for the game considered in [12]. Similarly, in the context of the method proposed in [12], the Shapley value (computed according to the specific choice of the characteristic function $c(V')$ detailed above) is used to rank joints in order of "importance," where the "most important" joint in a frame is one that has the largest Shapley value. In addition, the joint with the second-largest Shapley value is computed. Indeed, there may be two joints with the same largest rank (in this case, the "first" joint may be the one with the smallest index in a pre-defined labeling of joints). Looking also for the joint with the second-largest Shapley value is motivated by the fact that these two vertices could be connected by an edge in the weighted auxiliary graph (as occurred in one of the results presented in [12]). In that case, such an edge could be considered the "most important" edge in the graph for that specific frame.

Step 6. The last step of the method is the filtering of the computed Shapley values, to keep only the vertices automatically evaluated as the most important ones for some given number k of consecutive frames (selected, e.g., as $k = 51$, according to the implementation presented in [12]).

To conclude the presentation of the computational method developed in [12] for the automatic detection of the perceived OoM, Fig. 2 summarizes it by representing its conceptual architecture, clarifying the order of execution of its various steps, and reporting the main notation used in this section.

Finally, in the previous article [12], the proposed approach was validated using an online survey, in which participants were shown a series of triplets of videos displaying a skeletal representation of a dancer performing the same full-body expressive movement. Each video in the triplet had one highlighted joint (presented in random order): (1) joint with the maximum Shapley value; (2) joint with the maximum speed; (3) randomly chosen joint. The participants were asked to choose the video that better represented the evolution of the most relevant joint responsible for originating the dancer's movement. The results indicated that in most cases, the joint with the highest Shapley value was selected as the most relevant joint in the fragment. It was concluded that the detection of the joint with the maximum Shapley value is strongly correlated with the concept of perceived OoM in dance.

Fig. 2. Conceptual architecture of the computational method developed in [12] for the automatic detection of the perceived origin of movement (OoM).

4. The Annotated Movement Dataset

The dataset used in this work is made up of 36 unique video fragments acquired from multiple subjects, wearing infrared reflective markers. These fragments were recorded with the Qualisys MoCap system across multiple recording sessions and synchronized via SMPTE with two cameras (front and side views were available). The MoCap system included 13 high-resolution, high-precision infrared Oqus cameras and the Qualisys Track Manager (QTM) software, which extracted the 3D position vectors of the markers, from the corresponding 2D data collected by the Oqus cameras. In each fragment, a simple movement sequence was performed in which the OoM was clearly defined. Alongside each fragment were the MoCap position data and annotations for the recordings. These annotations included notes such as moving body parts, sequence of movement, and intended OoM.

This dataset was collected to create a repository of fragments of full-body movements characterized by a clearly perceived OoM. An example of such a fragment is one in which the subject performed a

movement originating at the shoulder with the rest of the body moving as a result. Such a fragment was accompanied by an annotation stating "left shoulder pulls the body making it turn." The set of annotations served as a baseline against which the effectiveness of a movement feature in automatically detecting the perceived OoM could be compared using the method described in Section 3.2.

4.1 The WhoLoDance Movement Dataset

This subset of the dataset was recorded in March 2016 at Casa Paganini (University of Genova), in the framework of the H2020-ICT-2015 EU Project WhoLoDance. Its sessions were largely characterized by multimodal recordings. As such, expressive aspects of movement (i.e., movement qualities) were the focus of such sessions. Some of these were, e.g., the (annotated) OoM, lightness, and fluidity. The subjects, who were professional dancers, prepared for their sessions by devising several exercises beforehand. The recordings consisted of contemporary dance movements and were carried out without music accompaniment, as this could have affected the way the dancers performed the movements.

For each trial of each recording session, manual annotations were kept by detailing the intended OoM, the rating of each trial—i.e., how accurate it was based on the brief—and the timestamp of the trial within the video of the session. Some examples of these manual annotations for the WhoLoDance recording sessions are shown in Table 1.

Table 1. Manual annotations (in *italics*), where the cells on the left denote the trial number and the leading joint(s), and the cells on the right describe if a task was performed and acquired correctly (e.g., "OK"), the rating of the task, and provide details of the movement and the timestamp of the task in the video

Trial 4	*OK. Rating: 5.*
Leading joint: left shoulder.	*Minutes: 00:09-00:16. Left shoulder pulls the body making it turn.*
Trial 5	*Not rated.*
Trial 6	*Task 1. Very separate. OK. Rating: 4.*
A sequence of movements originated from different parts of the body.	*Minutes: 00:10-00:22. Fingers - hands - right foot - head - hip - left arm - right arm.*

4.2 The Montpellier–UniGe Movement Dataset

This multimodal movement dataset was designed and recorded in July 2020 at Casa Paganini (University of Genova), as part of a collaboration between the Universities of Montpellier and Genova, in the context of the EU H2020 FET PROACTIVE EnTimeMent project. Its sessions were split into two parts, focusing respectively on individual actions and dual actions. Such actions are described in the next subsections. The objective was to detect the OoM in different goal-directed movements.

Action 1: Grasping

First, there was an individual grasping action, where participants faced an object located at shoulder length distance, 20° with respect to the external direction from the front of their dominant shoulder (see Fig. 3(a) for a visual representation of this action). The participants were asked to "spontaneously" reach for the object with their dominant hand and place it in front of the non-dominant side. The weight of the object was manipulated, as this could affect grasp configuration, posture, and even the OoM. As such, the participants were presented with two identical bottles, one filled with sand, the other empty, without knowing which was which. Then, the participants were asked to choose a bottle to start with, and the action was repeated for the other bottle.

Action 2: Throwing

The participants were asked to launch a ball using both hands. Employing both hands should facilitate the use of the whole body in a more controlled way. The participants were asked to avoid high-speed and

forceful throwing actions. They were also asked to stand in a non-symmetric position, e.g., with one foot a bit more forward than the other. This asymmetry allowed for a more "free" and natural movement.

This action was first done individually (Fig. 3(b)), where the participants launched the ball forward on their own. Additionally, a paired throwing action was also performed, in which the participants stood face-to-face at a fixed distance apart and launched the ball at each other (Fig. 3(c)). In the dual throwing action, the nature of the launch was manipulated in three ways: (i) fair launch; (ii) aggressive vs. defensive launch; (iii) cheating/pretense launch. In (i), the participants launched the ball to each other trying to facilitate the grasp by the other. In (ii), the sender launched the ball aggressively and angrily and the receiver caught the ball with a defensive behavior. Lastly, in (iii) the sender launched the ball in a way that aimed to reduce the successful grasp of the receiver, by pretending to throw the ball before launching it. Note that for analysis purposes, the sender and the receiver were considered separately, both in their position coordinates and in the visualization.

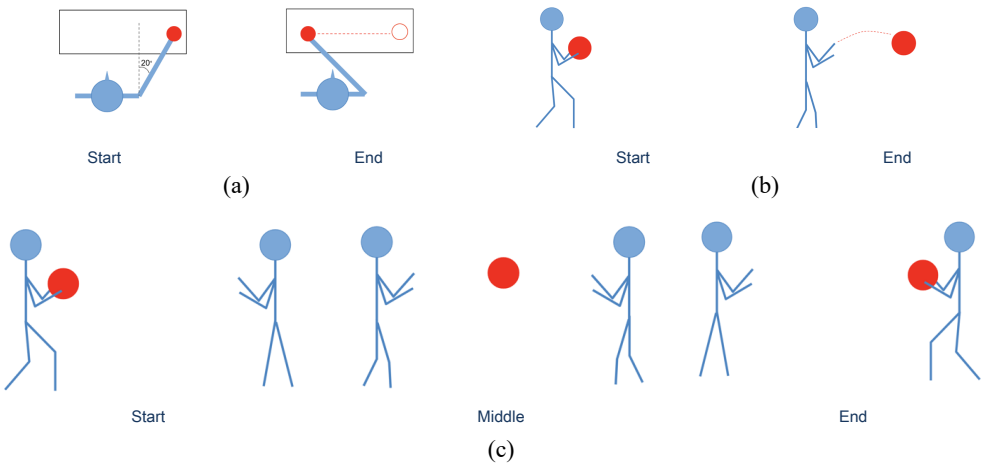


Fig. 3. Visual representations of actions in the Montpellier–UniGe movement dataset. (a) Experiment 1: individual grasping action (top view). (b) Experiment 2: Individual throwing action (side view). (c) Experiment 3: Dual throwing action (side view).

5. Marker Set

5.1 Full Marker Set

Because the two distinct subsets of the dataset were recorded with a 4-year time difference, there were some distinctions in their markers due to technological advances. Specifically, the full marker set for the WhoLoDance recordings consisted of 64 reflective markers, whereas the Montpellier–UniGe recordings used the new Qualisys sports marker set composed of 41 markers. However, the intuition behind the placement and clustering of the markers was the same for both. For simplicity, the next sections refer mainly to the case of the 64 marker set (which was acquired earlier).

5.2 Reduced Marker Set

The original dataset was reduced to a smaller dimension, i.e., it was transformed into a simplified dataset obtained by decreasing the number of markers. In practice, suitably constructed groups (clusters) of joints were reduced to a single joint (see Fig. 4 for details on the reduction from the set of 64 markers to the "reduced" set of 20 markers). This reduction was motivated by previous literature [29], which showed that a simplified skeletal structure can effectively convey relevant information on expressive

movements. Additionally, by combining multiple markers into a single cluster/joint, the risk of missing markers was reduced. It is also worth noting that the clusters considered in this reduction phase were determined a priori, differently from the clusters automatically detected by spectral clustering, as detailed in Section 3.2. The dataset reduction was performed by replacing the positions of the joints in each cluster with that of their center of mass. For example, the "left hand" cluster was associated with the center of mass of the left palm, thumb, middle, and pinky finger markers.

The coordinates of the center of mass of each cluster, for each frame of each fragment were calculated as follows. First, the same mass was assumed for each joint of the same cluster. The x, y, and z coordinates of the markers i belonging to each cluster (denoted as x_i, y_i, and z_i) were extracted. Then, for each of x, y, and z, its average on the cluster was computed. This was done, e.g., by summing all the x_i-coordinates together (and similarly for y_i and z_i), then dividing the result by the number of markers in the cluster.

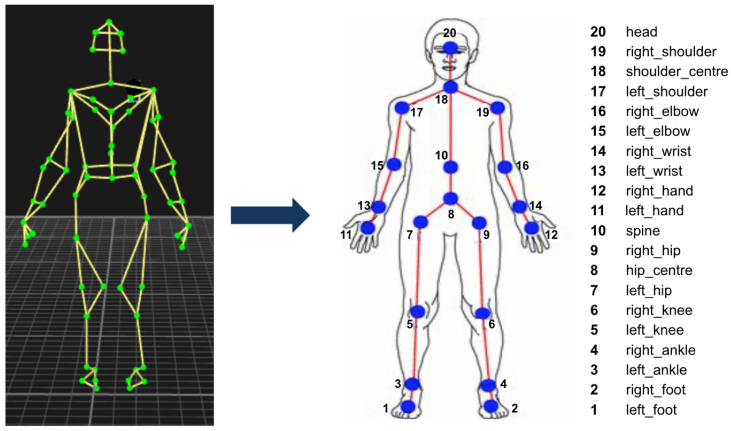


Fig. 4. Reduction of the skeletal structure with 64 markers to the "reduced" one with 20 markers. From the figure, it is possible to infer the mapping of vertices from the original structure to the "reduced" one. In this figure, "left" and "right" refer to the observer's viewpoint (not to the subject's viewpoint).

5.3 Mass Distribution Feature

As a result of the reduction of the marker set described in Section 5.2, some information may be lost in the mapping performed. Thus, an additional mass distribution (MD) feature was defined to determine the volume of information lost when moving from the full to the reduced marker set. The analysis presented later in this section showed that one could safely assume that each cluster in the original full marker set (which became a vertex in the skeletal structure associated with the reduced marker set, see Fig. 4) behaved approximately like a rigid body. For each cluster, the MD feature was defined as the root mean square distance of the markers in that cluster from their center of mass, i.e., $MD = \sqrt{\frac{\sum_{i=1}^{n} ||\boldsymbol{x}_i - \bar{\boldsymbol{x}}||^2}{n}}$, where n is the number of joints in the cluster, $\boldsymbol{x}_i$ is the position vector of each joint in the cluster (this information came from the MoCap data), and $\bar{\boldsymbol{x}}$ is the position vector of the center of mass of that cluster. The result of computing this feature was a scalar (in mm) for each cluster at each frame, whose purpose was to detect changes over time in the mass distribution of each feature. If no such notable changes were perceived, then joint rigidity could be assumed, i.e., joints defining each cluster in the full model were almost rigid with respect to one another. If this was the case, one could assume that the reduced model adequately approximated the full model. In case of an important change over time in the MD feature (see below for details), it was assumed that information was lost when moving from the full marker set to the

reduced one. In more detail, the coefficient of variation (or relative standard deviation) of the MD feature was calculated for each cluster to assess the dispersion of that feature over time. This coefficient is defined as the ratio between its empirical standard deviation and its empirical mean. To determine the coefficient of variation of the MD feature, clusters with less than two joints were excluded (i.e., according to Fig. 4, the left/right ankle and left/right shoulder clusters were not considered). A coefficient of variation larger than 0.05 was taken as an indication of the fact that information was lost in the process of model reduction.

The computation of the coefficient of variation (Table 2) showed that several clusters acted, indeed, as rigid bodies, because they were characterized by a small coefficient of variation. Other clusters, such as hand markers, had large coefficient values. It should be noted that measurement errors could have occurred during data recording, e.g., if markers on the extremities of the body shifted with subject movement. Thus, it was decided to ignore the coefficient of variation for clusters at the extremities of the body, assuming that they also acted as rigid bodies.

Table 2. Coefficient of variation (CV) of the mass distribution feature for each cluster of joints

Joint	CV	Joint	CV
Head	0.08	Hip center	0.06
Shoulder center	0.09	Spine	0.09
Left elbow	0.02	Left wrist	0.06
Right elbow	0.35	Right wrist	0.24
Left hand	0.12	Left foot	0.55
Right hand	0.13	Right foot	0.65
Left hip	0.27	Left knee	0.03
Right hip	0.35	Right knee	0.03

6. System Architecture

6.1 Pre-processing of the Data

As expected, some data were missing, producing NaN values in some frames. These "holes" were filled by interpolation based on rolling mean, rolling median, linear interpolation, spline, and polynomial interpolation. The most suitable method was then found for each data file, based on the R^2 value and the graph outputs.

Another important step in the pre-processing was outlier detection. In essence, the idea was to replace "problematic" data with NaNs when the magnitude of the acceleration of each joint was higher than that of typical human movements. In other words, a threshold equal to $5g$ was set, where $g = 9.8$ m/s^2 is the magnitude of the gravitational force. It is worth noting that this method also removed valid isolated points between two adjacent time intervals filled with NaNs. This created holes in the time series, which were then filled using splines. At this step, the cleaned data no longer contained high energy jumps due to outliers, allowing for simple filtering using a linear dynamical system. It is worth noting that the data were not filtered before, since linear filters cannot remove high-energy jumps without perturbing the signal significantly. Specifically, a second order lowpass Butterworth filter with an 8 Hz cutoff was applied to each cleaned position time series.

6.2 Computing Movement Features

Compared to the previous work [12], a larger number of movement features was considered in the present article (only speed was considered in [12]). In this way, it was possible to compare the results obtained using different features. In other words, the method detailed in Section 3.2 was used based on

different movement features for each joint (or for each cluster of joints, when the method was applied to the reduced dataset described in Section 5.2). Then their effectiveness in determining the OoM in selected fragments was estimated.

Speed: The first feature considered was speed (i.e., the magnitude of the tangential velocity $\boldsymbol{v}$), also used in the previous work [12], allowing for a direct comparison. Speed was calculated as the norm of the velocity vector (exploiting the *np.linalg.norm* method in Python). This choice was also motivated by the fact that speed is one of the most commonly used movement features when considering the perception of affective states, as was shown in a variety of psychological studies (see, e.g., [19]).

Tangential acceleration: The second movement feature considered was tangential acceleration (i.e., the rate of change of the magnitude of the tangential velocity $\boldsymbol{v}$). This feature was calculated with the *np.gradient* method in Python applied to the velocity vector.

Momentum: The third movement feature considered was the angular momentum around the center of mass of the whole body. The formula used to calculate this feature, for each joint (or cluster of joints) i and time t, is $\boldsymbol{L}_{i,MC} = \boldsymbol{r}_{i,MC} \times \boldsymbol{p}_{i,MC}$, where $\boldsymbol{r}_{i,MC}$ is its position vector at time t of i, and $\boldsymbol{p}_{i,MC} = m_i \boldsymbol{v}_{i,MC}$ is its linear momentum, where m_i is its mass and $\boldsymbol{v}_{i,MC} = \frac{d\boldsymbol{x}_{i,MC}}{dt}$ is its velocity. In this case, all these vectors $\boldsymbol{r}_{i,MC}$, $\boldsymbol{p}_{i,MC}$, $\boldsymbol{v}_{i,MC}$, and $\boldsymbol{x}_{i,MC}$ are relative to the center of mass of the whole body.

A movement in which parts of the body rotate with respect to different axes passing through the center of mass of the body has the same direction of angular momentum for different clusters if these clusters are rotating together. To determine clusters of joints with the same, or similar, direction of angular momentum $\boldsymbol{L}_{i,MC}$, the cosine similarity was taken as a measure of similarity of that feature (to be used in the method described in Section 3.2). In other words, for each physical or non-physical edge, the cosine similarity was computed between the angular momentum feature computed at the two associated vertices, e.g., at the hand and the respective wrist. Thus, the cosine similarity was used to compare the directions of the angular momenta associated with the two vertices. It is important to note that the cosine similarity measure ranges from -1 to 1. Applying the method detailed in Section 3.2 to a generic movement feature required non-negative similarity values to define the weights of the edges of the graph G. This was easily solved by incrementing the cosine similarity by 1, altering its range to [0, 2], with 2 being the case where pairs of vertices had the same direction of the angular momentum feature.

6.3 Calculation of the Shapley Values

At the end of the step described in Section 6.2, it was possible to obtain, for each frame of each fragment, a variable that contained the movement features of all the joints. The next step was, for each of these features, to calculate the set of Shapley values (one for each player). In essence, to compute the Shapley values, first, the data with the values calculated for the three movement features were loaded. Then, the reduced skeletal structure was constructed as described in Fig. 4 (the figure also reports the indexing of the joints).

In particular, the following vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ were used to encode the connections between pairs of joints in the reduced model: $\boldsymbol{n}_1 = [20\ 18\ 17\ 15\ 13\ 18\ 19\ 16\ 14\ 18\ 10\ 8\ 7\ 5\ 3\ 8\ 9\ 6\ 4]$, $\boldsymbol{n}_2 = [18\ 17\ 15\ 13\ 11\ 19\ 16\ 14\ 12\ 10\ 8\ 7\ 5\ 3\ 1\ 9\ 6\ 4\ 2]$. For instance, referring to the first components of $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$, one gets that (20, 18) = (head, shoulder center) is the first edge of the reduced skeletal structure, as the first component of $\boldsymbol{n}_1$ is 20 (head), and the first component of $\boldsymbol{n}_2$ is 18 (shoulder center).

Subsequently, spectral clustering was applied to the weighted graph G constructed in the way described in Section 3.2. At this point, it was possible to compute the Shapley values of all 20 joints for all the different movement features for each frame, following the method reported therein. The most relevant joint—i.e., the one with the largest rank coming from the Shapley values—was automatically determined. The output of this algorithm was a table in which one index refers to one of the frames in a fragment, and the other index refers to one of the top 10 largest normalized Shapley values. The normalization was performed by dividing each Shapley value by the largest Shapley value in the same frame, and was

intended to facilitate the comparison of the results obtained in different frames However, since we were not interested in analyzing all 10 joints, we focused only on the top 10 joints.

Fig. 5 reports, for three frames of a fragment, the joints associated with the 10 largest normalized Shapley values, and the corresponding normalized values. The fragment considered refers to one of the individual throwing actions in which the subject used her knees to begin the action of throwing the ball (Fig. 3(b)). As such, the OoMs were the left knee and the right knee. Fig. 5 shows that the top three normalized Shapley values (colored in purple) in all three frames were either knee joints or ankle joints, which are physically connected to the knee joints. Moreover, such joints had a normalized Shapley value equal to 1.0 in most cases. This shows the effectiveness of estimating the OoM through Shapley values, for a TU model whose utility function is based on movement features (here in particular, speed was used as the movement feature of the method).

Frame 1.name	**Frame 1**.value	**Frame 2**.name	**Frame 2**.value	**Frame 3**.name	**Frame 3**.value
right_ankle	1.0	left_knee	1.0	right_ankle	1.0
right_knee	1.0	right_ankle	0.97	right_knee	1.0
shoulder_center	0.8	right_knee	0.97	left_knee	0.85
head	0.8	left_ankle	0.75	left_ankle	0.64
left_elbow	0.76	left_elbow	0.53	shoulder_center	0.53
right_elbow	0.46	left_shoulder	0.27	head	0.49
left_wrist	0.42	left_wrist	0.26	left_elbow	0.25
left_knee	0.37	left_hip	0.25	left_shoulder	0.25
left_hip	0.37	right_elbow	0.21	left_hip	0.21
right_shoulder	0.34	right_shoulder	0.21	right_elbow	0.15

Fig. 5. Normalized Shapley values of a fragment, calculated using speed as the movement feature.

7. Validation

7.1 Comparison Framework

As mentioned before, the available dataset consisted of videos accompanied by expert annotations about the specific movements performed. These primary expert annotations comprised the acted OoM for each fragment and could be used, for a given fragment of a video, to measure the accuracy (in percentage) of a movement feature in estimating the perceived OoM. In essence, these manual expert annotations acted as a ground truth against which one could compare the estimated OoM using a certain movement feature. Three methods of comparison were used. First, the ground truth was compared to a choice of joint at random, creating a minimal baseline for the other methods. The hypothesis under this baseline case is that its accuracy score would be approximately 5% since 20 joints were considered in total in the reduced model. Second, the ground truth was compared to the joint with the largest normalized Shapley value. Last, the ground truth was compared to the joints with the two largest normalized Shapley values, to get a more "relaxed" comparison. In this way, it was possible to obtain accuracy scores by restricting to frames that satisfied the additional condition that the absolute value of the difference between the first- and the second-largest normalized Shapley values was less than 0.05. Note that for the first two comparisons, the accuracy was computed with respect to the total number of frames, while for the third comparison, it was computed with respect to the number of frames in which the two largest normalized Shapley values were (almost) equal. As an example, Table 3 shows the results of these comparisons for a specific fragment in the available dataset, based on all three movement features.

As anticipated, the first method of comparison with the random joint choice resulted in a much lower accuracy score than the other two, close to 5% as was hypothesized for all the movement features. It is evident that speed yielded the highest accuracy scores for all the comparisons, particularly for the third one. From these results obtained for all fragments, it was possible to gain insight into the performance of

each movement feature. However, there was no information about the relevance of certain frames of a fragment. It is also important to note that the ground truth, which consisted of manual annotations, was generated by one expert only, so it was bound to produce preliminary results.

Table 3. Summary of accuracy results for a single fragment

Fragment t_028.3	**Accuracy (%)**		
	Speed	**Acceleration**	**Angular momentum**
Random choice	5.81	5.44	5.43
First-largest normalized Shapley value	19.42	13.07	19.16
First- and second-largest normalized Shapley values	80.36	51.18	50.89

7.2 Online Annotation Tool

All 36 fragments that made up the dataset were uploaded to a server. In total, 127 participants annotated the fragments. The online annotation tool was essential to the completion and validation of the research reported in this article, as allowed the extraction of individual ground truths, i.e., individual perceived OoMs, which could be directly compared to the acted OoMs for each fragment, provided by the manual expert annotations.

At the top of the page shown to the participant was the video fragment, which could be re-played by each participant as many times as needed. The participants were given a list of the 20 joints from which they could choose up to two as their perception of the OoM. At the bottom of the page, there was a five-level Likert scale, on which the participants could indicate the level of confidence in their answers, i.e., how sure they were of their choice of the OoM. It is worth noting that participants viewed only 18 fragments per session. Each fragment was just a few seconds long, so one session would take a maximum of about 10–15 minutes to complete. A randomization algorithm was also implemented, to ensure that the fragments were not viewed in any particular order and were interchanged between sessions to guarantee that all the fragments were viewed the same number of times.

8. Results

8.1 Accuracy Scores

8.1.1 Accuracy scores of movement features

This section describes the results obtained regarding the accuracy scores of the movement features in determining the OoM. Table 4 details the maximum accuracy scores (with respect to the fragments) for the three movement features across the second and third comparisons described in Section 7.1.

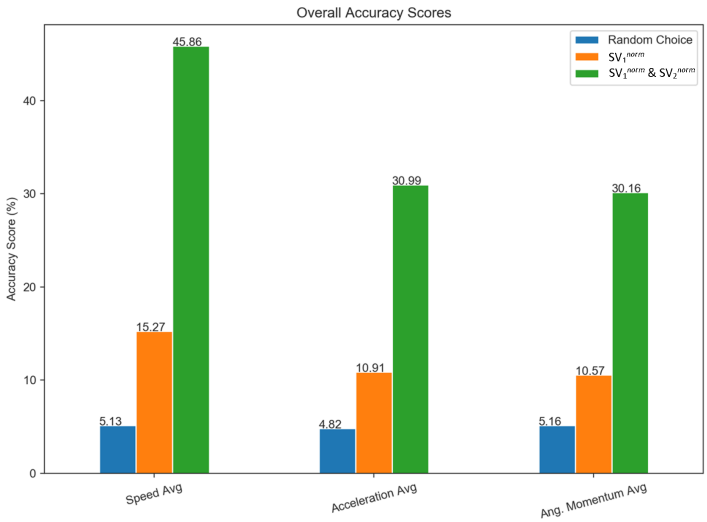
Table 4. Maximum accuracy scores (with respect to the fragments) for the second and third comparisons across all movement features

	Maximum accuracy (%)	
	Comparison 2	**Comparison 3**
Speed	59.13	91.21
Acceleration	29.95	54.74
Angular momentum	32.25	62.64

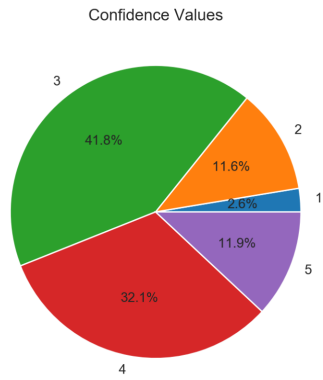
As a reminder, the second comparison looked at the intersection between the ground truth and the joint with the first-largest normalized Shapley value, and the third comparison looked at the intersection

between the ground truth and the two joints associated with the first- and the second-largest normalized Shapley values SV_1^{norm} and SV_2^{norm}, for the frames where $|SV_1^{norm} - SV_2^{norm}| < 0.05$.

The results reported in Table 4 show an accuracy score of 91.21% for speed in the third comparison, representing an almost complete intersection between the acted OoM and the OoM as it was automatically perceived by the method. Overall, these maximum accuracy scores are encouraging and could serve as a stepping stone for further research. An overview of the average accuracy scores (with respect to all the fragments) for each comparison framework and each movement feature is reported in Fig. 6(a). From that figure it is possible to see that, mainly for the orange and green bars (second and third comparisons), speed outperformed on average the other movement features, which is consistent with the results in Table 4. Additionally, as hypothesized, the accuracy in the first comparison was about 5%, illustrating how the random estimation of the OoM did not produce satisfying results.



(a)



(b)

Fig. 6. (a) Average accuracy scores for speed, tangential acceleration, and angular momentum, across the following three comparison frameworks. In blue: comparison of the ground truth against a random generation of joints. In orange: comparison of the ground truth against the joint associated with the first-largest normalized Shapley Value. In green: comparison of the ground truth against the two joints associated with the first- and the second-largest normalized Shapley values SV_1^{norm} and SV_2^{norm}, for the frames in which $|SV_1^{norm} - SV_2^{norm}| < 0.05$. (b) Pie chart of confidence values across all the fragments.

These results confirm those obtained in [12] on a different dataset, where speed was the movement feature provided as input to the method. It was found that participants selected the joint associated with the largest Shapley value more frequently than the one associated with the largest speed. It was also established in [12] that the selection frequency of the joint with the largest speed increased as the level of expertise of the participants decreased (i.e., moving from professional dancers to non-dancers).

8.1.2 Comparing perceived vs. acted origin of movement

Another aspect of accuracy that was worth investigating is the intersection between the acted OoM and the perceived OoM for each fragment, the latter being obtained using the online annotation tool. This comparison was done as follows. First, the number of occurrences of the most common selection per fragment was determined for all participants. For example, if for one fragment, the left hand was selected 40 times and the left elbow was selected 35 times, then the left hand was chosen as the most common answer. Then, this selection was compared with the acted OoM coming from the manual annotations made in each fragment.

In summary, from these comparisons it was possible to extract an accuracy score for the intersection between acted and perceived OoM, i.e., it was computed how similar the manual annotations were to the online annotations made by the participants. Table 5 reports that the mean accuracy score (with respect to the fragments) of this intersection was 31.1%, with the maximum intersection score equal to 53.41% for one fragment. Overall, these values were satisfactory, considering the range of selection of choices available to the participants—i.e., the 20 joints—and the number of fragments in the analysis.

Table 5. Accuracy of the intersection between manual annotations (ground truth/acted OoMs) and most common participants' online annotations (perceived OoMs)

	Acted = perceived OoM
Mean	31.10%
Max	53.41%

OoM = origin of movement.

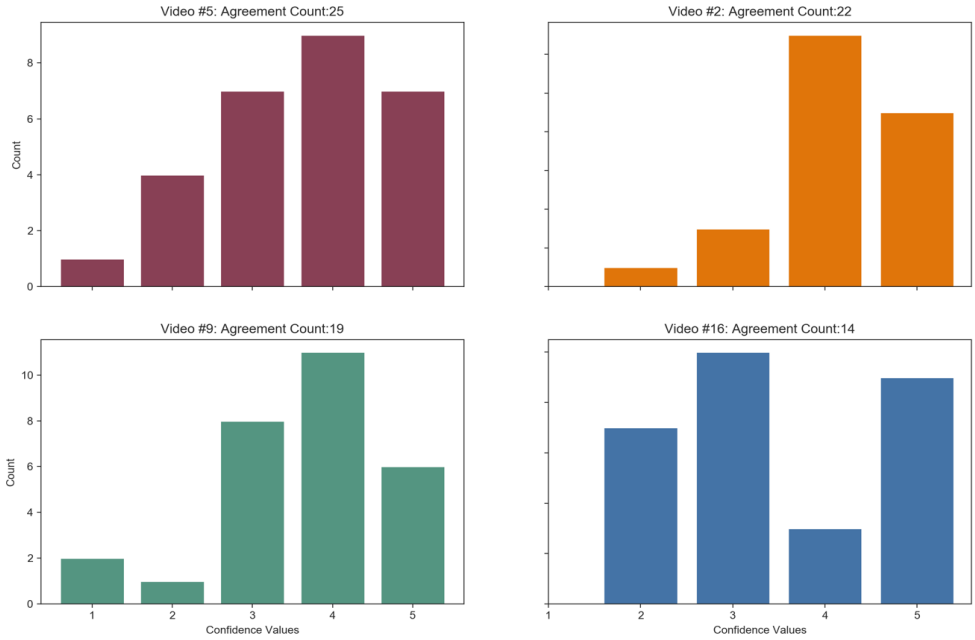
8.2 Confidence Values

It is worth observing that, as shown in Fig. 6(b), participants tended to be quite confident in their selections of joints, with the majority of them selecting 3 or 4 on the confidence scale. In particular, over 40% of the participants were neutral in their selections (corresponding to 3 on the Likert scale), while approximately 55% of them were very confident in their selections (corresponding to 4 and 5 on the Likert scale). Only a very small percentage, 14.2%, reported an overall lack of confidence and uncertainty in their answers.

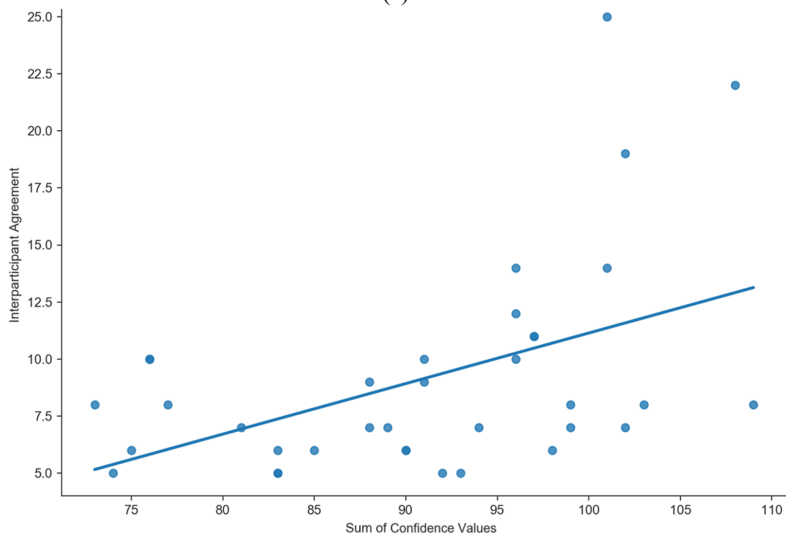
8.3 Inter-participant Agreement

Another important piece of information extracted from the analysis of the online validation survey was the inter-participant agreement, i.e., the largest number of participants who selected the same joint as the OoM in a given fragment. The presence of a strong inter-participant agreement suggests that there was a clear perception of the OoM for that fragment. Another important aspect concerns the fragments with a high inter-participant agreement, for which the confidence values were also on the higher side of the Likert scale. A combination of high inter-participant agreement and high confidence values is an indicator that a fragment has a clearly perceived OoM. Hence, such a fragment seems to be relevant for further analysis. Fig. 7(a) represents the four fragments with the highest inter-participant agreement scores. For each of those fragments, a bar chart was created, depicting the distribution of the participants' confidence values. These turned out to be high overall, supporting the relevance of these fragments for further movement analysis. To address the relationship between confidence values and inter-participant agreement,

a Spearman correlation test was performed on the data in Fig. 7(b), yielding a positive correlation (0.449). While this value was not high, it confirmed the relevance of fragments with the highest inter-participant agreement scores.



(a)



(b)

Fig. 7. (a) Bar charts showing the distribution of confidence values from the four fragments with the highest inter-participant agreement scores (video numbers: 2, 5, 9, and 16). (b) Scatter plot of the sum of confidence values per fragment against the inter-participant agreement per fragment.

8.4 Fragments with No Clear Perceived Origin of Movement

This section concludes the presentation of the results achieved by the method by analyzing the effect of the presence of fragments with no clear perceived OoM on the online annotation process by the participants. Indeed, for a better validation of the method, the dataset included such a kind of fragment.

There were two such fragments. The first consisted of a subject performing calibration movements (based on a T-pose) for the MoCap cameras. The second fragment with no clear perceived OoM represented a subject randomly walking around the stage between two successive trials. Such a fragment was taken from a recording in which a subject performed the solo throwing action, then walked across the screen to pick up the ball after throwing it, and finally went back to the initial position.

As expected, the results of the online validation survey indicated that there was no clear consensus on the perceived OoM in these two cases. Fig. 8 illustrates, for the two fragments, the joints selected by the participants as the perceived OoM. Such selections are highlighted in green. For the first fragment, Fig. 8(a) shows that the participants' joint selection ranged from the top to the bottom of the body, i.e., from the head to the right foot. For the second fragment, Fig. 8(b) shows that there was quite an equal distribution of selections between the left and the right side of the body, which makes sense since walking requires both sides of the body. However, both cases clearly exhibited a lack of agreement among the participants, further demonstrating the importance of the online validation survey.

(a) (b)

Fig. 8. Body model with the joints (green) selected as the perceived OoMs for two fragments with no clear perceived OoM. (a) First fragment: T-Pose. (b) Second fragment: random walking.

9. Conclusion and Future Work

A computational method for the automatic detection of the perceived OoM was further developed. This research is relevant because such detection can allow one to interpret affective content and social signals and can have applications in cognitive/motor rehabilitation, among others.

In general, the results of the analysis demonstrated the effectiveness of the method. Still, there are several possible future extensions of this work, which could improve it and yield robust results, overcoming the current limitations of the method. First, its online implementation (based, e.g., on a real-time pre-processing of its inputs) would allow the method to be applied not only in laboratory research but also in several real-world contexts: e.g., as a novel application of artificial intelligence to healthcare [30], for tasks related to cognitive/motor rehabilitation. However, its application to healthcare is also expected to raise issues related to security and privacy, which could be among the possible subjects of further research (see, e.g., [31–34] for possible approaches related to blockchain technology). Second, more detailed skeleton structures could be considered by the method. For instance, one could use a sequence of skeleton structures corresponding to nested sets of markers. This would allow comparisons of the outcomes of the method (in terms of the automatically detected origin of movement) obtained by using two or more skeleton models at different levels of detail. Alternatively, one could perform the analysis directly on the most detailed model and replace the clustering with hierarchical clustering [35]. Third, a larger set of features calculated from movements could be considered. To address this, one could construct a vector of m features (with $m \geq 2$) and search for the most relevant subset of them with

cardinality at most $\bar{m} < m$. However, this could be very computationally demanding, as one would have to consider several such subsets and provide a suitable weight to each feature to get a (scalar) similarity measure between vertices of the graph G. Furthermore, considering biomechanics of human movement could provide further insights into this research. Indeed, biomechanics constrains the way one moves as well as the way one perceives movements. In our game-theoretic model, biomechanical properties could be further considered as constraints, with the expectation that such constraints would restrict the range of possible values assumed by the Shapley values of different joints. Moreover, the origin of movement could be investigated at different temporal scales. For example, one could look at a smaller temporal scale at the very first phase of the movement, and then also at a longer temporal scale. This would allow one to analyze the origin of movement at a higher level. Thus, multiple temporal scales would be useful to perform different levels of analysis and make it feasible to investigate human behavior in terms of the origin of movement based on sophisticated approaches [36]. Future research could also consist of applying state-of-the-art signal processing algorithms as an additional pre-preprocessing step (e.g., the one presented in [37]). This pre-processing could be based on numerical discretization of derivatives (see also [38, 39]), or image processing algorithms [40, 41], aimed at denoising the MoCap data (i.e., the frame-by-frame position vectors of the markers). Such a pre-preprocessing step would be useful, e.g., in the case of missing/unlabeled markers (in some frames) or in the case of noisy position measures.

Author's Contributions

Conceptualization, BB, AC, GG, OM, DM, MS; Funding acquisition, AC, GG; Investigation and methodology, BB, AC, GG, OM, DM, MS; Writing of the original draft, OM; Writing of the review and editing, BB, AC, GG, OM, DM, MS; Software, OM; Validation, OM; Visualization, OM.

Funding

The work received support from the 4-year European Project no. 824160 FET PROACTIVE EnTimeMent. G. Gnecco was supported in part by the Galileo 2021 project no. G21_89 "Automatic Movement Analysis Techniques for Applications in Cognitive/Motor Rehabilitation" between Italy and France, by the project "THE – Tuscany Health Ecosystem" (CUP: D63C22000400001), funded by the European Union - Next Generation EU program, in the context of the Italian National Recovery and Resilience Plan, Investment 1.5: Ecosystems of Innovation, and by the PRIN PNRR 2022 project MOTUS – Automated Analysis and Prediction of Human Movement Qualities" (CUP: D53D23017470001), funded by the European Union – Next Generation EU program. M. Sanguineti was supported in part by the National Research Council of Italy (CNR), where he is a Research Associate at INM – Institute of Marine Engineering. Part of the research was conducted during the stay of M. Sanguineti at IMT School for Advanced Studies Lucca, Italy, as a Visiting Professor. G. Gnecco and M. Sanguineti are members of GNAMPA – National Group for Mathematical Analysis, Probability and their Applications of INdAM – National Institute of Higher Mathematics. G. Gnecco dedicates the work to the memory of his mother Rosanna Merlini.

Competing Interests

The authors declare that they have no competing interests.

References

[1] M. Kazemitabar, S. P. Lajoie, and T. Doleck, "Analysis of emotion regulation using posture, voice, and attention: a qualitative case study," *Computers and Education Open*, vol. 2, article no. 1100030, 2021. https://doi.org/10.1016/j.caeo.2021.100030

[2] E. Elkjær, M. B. Mikkelsen, J. Michalak, D. S. Mennin, and M. S. O'Toole, "Expansive and contractive postures and movement: a systematic review and meta-analysis of the effect of motor displays on affective and behavioral responses," *Perspectives on Psychological Science*, vol. 17, no. 1, pp. 276-304, 2022. https://doi.org/10.1177/1745691620919358

[3] M. A. Mahfoudi, A. Meyer, T. Gaudin, A. Buendia, and S. Bouakaz, "Emotion expression in human body posture and movement: a survey on intelligible motion factors, quantification and validation," *IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 2697-2721, 2023. https://doi.org/10.1109/TAFFC.2022.3226252

[4] M. Behnke, N. Bianchi-Berthouze, and L. D. Kaczmarek, "Head movement differs for positive and negative emotions in video recordings of sitting individuals," *Scientific Reports*, vol. 11, article no. 7405, 2021. https://doi.org/10.1038/s41598-021-86841-8

[5] J. F. Christensen, R. T. Azevedo, and M. Tsakiris, "Emotion matters: different psychophysiological responses to expressive and non-expressive full-body movements," Acta Psychologica, vol. 212, article no. 103215, 2021. https://doi.org/10.1016/j.actpsy.2020.103215

[6] M. M. N. Bienkiewicz, A. P. Smykovskyi, T. Olugbade, S. Janaqi, A. Camurri, N. Bianchi-Berthouze, M. Bjorkman, and B. Bardy, "Bridging the gap between emotion and joint action," *Neuroscience & Biobehavioral Review*, vol. 131, pp. 806-833, 2021. https://doi.org/10.1016/j.neubiorev.2021.08.014

[7] X. Xi, Q. Tao, J. Li, W. Kong, Y. B. Zhao, H. Wang, and J. Wang, "Emotion-movement relationship: a study using functional brain network and cortico-muscular coupling," *Journal of Neuroscience Methods*, vol. 362, article no. 109320, 2021. https://doi.org/10.1016/j.jneumeth.2021.109320

[8] X. Sun, K. Su, and C. Fan, "VFL: a deep learning-based framework for classifying walking gaits into emotions," *Neurocomputing*, vol. 473, pp. 1-13, 2022. https://doi.org/10.1016/j.neucom.2021.12.007

[9] K. Hook, *Designing with the Body: Somaesthetic Interaction Design*. Cambridge, MA: MIT Press, 2018.

[10] K. K. A. Bakhti, D. Mottet, N. Schweighofer, J. Froger, and I. Laffont, "Proximal arm non-use when reaching after a stroke," *Neuroscience Letters*, vol. 657, pp. 91-96, 2017. https://doi.org/10.1016/j.neulet.2017.07.055

[11] K. K. A. Bakhti, I. Laffont, M. Muthalib, J. Froger, and D. Mottet, "Kinect-based assessment of proximal arm non-use after a stroke," *Journal of NeuroEngineering and Rehabilitation*, vol. 15, article no. 104, 2018. https://doi.org/10.1186/s12984-018-0451-2

[12] K. Kolykhalova, G. Gnecco, M. Sanguineti, G. Volpe, and A. Camurri, "Automated analysis of the origin of movement: an approach based on cooperative games on graphs," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 6, pp. 550-560, 2020. https://doi.org/10.1109/THMS.2020.3016085

[13] M. Maschler, S. Zamir, and E. Solan, *Game Theory*, 2nd ed. New York, NY: Cambridge University Press, 2020.

[14] O. Matthiopoulou, B. Bardy, G. Gnecco, D. Mottet, M. Sanguineti, and A. Camurri, "A computational method to automatically detect the perceived origin of full-body human movement and its propagation," in *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI)*, Virtual Event, The Netherlands, 2020, pp. 449-453. https://doi.org/10.1145/3395035.3425971

[15] A. Camurri and G. Volpe, "The intersection of art and technology," *IEEE MultiMedia*, vol. 23, no. 1, pp. 10-17, 2016. https://doi.org/10.1109/MMUL.2016.13

[16] J. Laroche, A. Tomassini, G. Volpe, A. Camurri, L. Fadiga, and A. D'Ausilio, "Interpersonal sensorimotor communication shapes intrapersonal coordination in a musical ensemble," *Frontiers in Human Neuroscience*, vol. 16, article no. 899676, 2022. https://doi.org/10.3389/fnhum.2022.899676

[17] M. Karg, A. A. Samadani, R. Gorbet, K. Kuhnlenz, J. Hoey, and D. Kulic, "Body movements for affective expression: a survey of automatic recognition and generation," *IEEE Transactions on Affective Computing*, vol. 4, no. 4, pp. 341-359, 2013. https://doi.org/10.1109/T-AFFC.2013.29

[18] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15-33, 2013. https://doi.org/10.1109/T-AFFC.2012.16

[19] B. De Gelder, "Why bodies? Twelve reasons for including bodily expressions in affective neuroscience," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3475-3484, 2009. https://doi.org/10.1098/rstb.2009.0190

[20] A. Camurri, I. Lagerlof, and G. Volpe, "Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 213–225, 2003. https://doi.org/10.1016/S1071-5819(03)00050-8

[21] M. J. Vaessen, E. Abassi, M. Mancini, A. Camurri, and B. De Gelder, "Computational feature analysis of body movements reveals hierarchical brain organization," *Cerebral Cortex*, vol. 29, no. 8, pp. 3551-3560, 2018. https://doi.org/10.1093/cercor/bhy228

[22] N. Dounskaia, "Control of human limb movements: the leading joint hypothesis and its practical applications," *Exercise and Sport Sciences Reviews*, vol. 38, no. 4, pp. 201-208, 2010. https://doi.org/10.1097/JES.0b013e3181f45194

[23] D. A. Simovici, *Clustering: Theoretical and Practical Aspects*. Singapore: World Scientific Publishing, 2021.

[24] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, pp. 395-416, 2007. https://doi.org/10.1007/s11222-007-9033-z

[25] Y. Hadas, G. Gnecco, and M. Sanguineti, "An approach to transportation network analysis via transferable utility games," *Transportation Research Part B: Methodological*, vol. 105, pp. 120-143, 2017. https://doi.org/10.1016/j.trb.2017.08.029

[26] G. Gnecco, Y. Hadas, and M. Sanguineti, "Public transport transfers assessment via transferable utility games and Shapley value approximation," *Transportmetrica A: Transport Science*, vol. 17, no. 4, pp. 540-565, 2021. https://doi.org/10.1080/23249935.2020.1799112

[27] M. Passacantando, G. Gnecco, Y. Hadas, and M. Sanguineti, "Braess' paradox: a cooperative game-theoretic point of view," *Networks*, vol. 78, no. 3, pp. 264-283, 2021. https://doi.org/10.1002/net.22018

[28] G. Gnecco, Y. Hadas, and M. Sanguineti, "A game theoretic approach for reliability evaluation of public transportation transfers with stochastic travel and waiting times," *Euro Journal on Transportation and Logistics*, vol. 11, article no. 100090, 2022. https://doi.org/10.1016/j.ejtl.2022.100090

[29] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Toward a minimal representation of affective gestures," *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 106-118, 2011. https://doi.org/10.1109/T-AFFC.2011.7

[30] G. Dicuonzo, F. Donofrio, A. Fusco, and M. Shini, "Healthcare system: moving forward with artificial intelligence," *Technovation*, vol. 120, article no. 102510, 2023. https://doi.org/10.1016/j.technovation.2022.102510

[31] A. A. Khan, A. A. Wagan, A. A. Laghari, A. R. Gilal, I. A. Aziz, and B. A. Talpur, "BIoMT: a state-of-the-art consortium serverless network architecture for healthcare system using blockchain smart contracts," *IEEE Access*, vol. 10, pp. 78887-78898, 2022. https://doi.org/10.1109/ACCESS.2022.3194195

[32] P. Kumar, R. Kumar, G. P. Gupta, R. Tripathi, A. Jolfaei, and A. N. Islam, "A blockchain-orchestrated deep learning approach for secure data transmission in IoT-enabled healthcare system," *Journal of Parallel and Distributed Computing*, vol. 172, pp. 69-83, 2023. https://doi.org/10.1016/j.jpdc.2022.10.002

[33] A. A. Khan, A. A. Laghari, Z. A. Shaikh, Z. Dacko-Pikiewicz, and S. Kot, "Internet of Things (IoT) security with blockchain technology: a state-of-the-art review," *IEEE Access*, vol. 10, pp. 122679-122695, 2022. https://doi.org/10.1109/ACCESS.2022.3223370

[34] P. K. Ghosh, A. Chakraborty, M. Hasan, K. Rashid, and A. H. Siddique, "Blockchain application in health-care systems: a review," *Systems*, vol. 11, no. 1, article no. 38, 2023. https://doi.org/10.3390/systems11010038

[35] R. Metulini, G. Gnecco, F. Biancalani, and M. Riccaboni, "Hierarchical clustering and matrix completion for the reconstruction of world input–output tables," *AStA Advances in Statistical Analysis*, vol. 107, no. 3, pp. 575-620, 2023. https://doi.org/10.1007/s10182-022-00448-6

[36] H. El-Hussieny, A. A. Abouelsoud, S. F. Assal, and S. M. Megahed, "Adaptive learning of human motor behaviors: an evolving inverse optimal control approach," *Engineering Applications of Artificial Intelligence*, vol. 50, pp. 115-124, 2016. https://doi.org/10.1016/j.engappai.2016.01.024

[37] A. Bacigalupo, M. L. De Bellis, G. Gnecco, and F. Nutarelli, "On dispersion curve coloring for mechanical metafilters," *Scientific Reports*, vol. 12, article no. 20019, 2022. https://doi.org/10.1038/s41598-022-23491-4

[38] O. A. Arqub and Z. Abo-Hammour, "Numerical solution of systems of second-order boundary value problems using continuous genetic algorithm," *Information Sciences*, vol. 279, pp. 396-415, 2014. https://doi.org/10.1016/j.ins.2014.03.128

[39] Z. Abo-Hammour, O. Abu Arqub, S. Momani, and N. Shawagfeh, "Optimization solution of Troesch's and Bratu's problems of ordinary type using novel continuous genetic algorithm," *Discrete Dynamics in Nature and Society*, vol. 2014, article no. 401696, 2014. https://doi.org/10.1155/2014/401696

[40] Y. Chen, R. Xia, K. Zou, and K. Yang, "FFTI: image inpainting algorithm via features fusion and two-steps inpainting," *Journal of Visual Communication and Image Representation*, vol. 91, article no. 103776, 2023. https://doi.org/10.1016/j.jvcir.2023.103776

[41] Y. Chen, R. Xia, K. Yang, and K. Zou, "MFFN: image super-resolution via multi-level features fusion network," *The Visual Computer*, vol. 40, pp. 489-504, 2024. https://doi.org/10.1007/s00371-023-02795-0