



**HAL**  
open science

## Expressive motions recognition and analysis with learning and statistical methods

Insaf Ajili, Zahra Ramezanpanah, Malik Mallem, Jean-Yves Didier

► **To cite this version:**

Insaf Ajili, Zahra Ramezanpanah, Malik Mallem, Jean-Yves Didier. Expressive motions recognition and analysis with learning and statistical methods. *Multimedia Tools and Applications*, 2019, 78 (12), pp.16575–16600. 10.1007/s11042-018-6893-5 . hal-01971017

**HAL Id: hal-01971017**

**<https://hal.science/hal-01971017v1>**

Submitted on 6 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Expressive motions recognition and analysis with learning and statistical methods

Insaf Ajili · Zahra Ramezanpanah · Malik Malleem · Jean-Yves Didier

the date of receipt and acceptance should be inserted later

**Abstract** This paper proposes to recognize and analyze expressive gestures using a descriptive motion language, the Laban Movement Analysis (LMA) method. We extract body features based on LMA factors which describe both quantitative and qualitative aspects of human movement. In the direction of our study, a dataset of 5 gestures performed with 4 emotions is created using the motion capture Xsens. We used two different approaches for emotions analysis and recognition. The first one is based on a machine learning method, the Random Decision Forest. The second approach is based on the human's perception. We derived the most important features for each expressed emotion using the same methods, the RDF, and the human's ratings. A comparison between RDF and observers classifiers was made in the discussion section.

**Keywords** Expressive motions recognition, Laban Movement Analysis, Random Decision Forest, human perception, Features importance

## 1 Introduction

Body language often has a great impact in transmitting messages to the listener by providing social meaningfulness to daily human-human interactions. Among the various type of body languages, facial expressions, body posture, and gestures

---

I. Ajili  
IBISC, Univ Evry, Université Paris-Saclay, 91025, Evry, France  
Tel: 01 69 47 75 51  
Fax: 01.69.47.06.14  
E-mail: insaf.bouzayaniajili@univ-evry.fr

Z. Ramezanpanah  
E-mail: zahra.ramezanpanah@gmail.com

M. Malleem  
E-mail: abdelmalik.malleem@univ-evry.fr

J.Y. Didier  
E-mail: jeanyves.didier@univ-evry.fr

are very important to convey feelings and contribute to the overall effectiveness of communication. The expressive quality of movement presents the way in which a gesture is performed. It is an important factor for transmitting high-level pieces of information during human being communication (e.g., emotions, moods). Generally, the state of the art research has focused on the decoding of emotions from facial expressions and vocal cues. A smile or raised inner eyebrows can convey happiness and amusement. Eyebrow lower and chin raise can characterize sad emotion. More than 95% of the literature in the field of identification of human emotions has used the facial expressions as a modality for emotion perception [1]. Besides that, few researchers have studied how body motions can provide reliable characteristics of emotions across the lifespan. In this work, we aim to develop an automated expressive motion recognition system which characterizes human behaviors (motions and emotions) based on his body movement without taking into consideration the facial expressions. In daily activities, happiness is always illustrated by jumping, sadness by bending, anger by punching, etc [2]. But on the other hand, the same motions performed with different intentions can also give us pieces of information about human psychological states. According to psychology research, a prerequisite to analyzing the human behavior is to perceive which motion characteristics lead to express a specific type of emotions. For example, sadness was always characterized by smooth and slow movements. However, Happiness was associated with faster, and more expanded movements [3]. However, human physiology and the limitation in the physical behavior of human movement cause the emotion recognition to be a difficult task.

The main purpose of our work is to extract a robust motion descriptor in order to characterize and recognize human emotions performed with different motions. On the other hand, we interpret the relationship between motions parameters and emotions. Many motion features were proposed in the literature with different combinations. However, there is always a lack of significant formalization in the direction of evaluating their perceptual effect. To address the above issues, we used the Laban Movement Analysis (LMA) technique, which is a language for describing and understanding human movement. This method captures both quantitative and qualitative aspects of the movement. Based on its four factors (Body, Space, Shape, and Effort), the LMA helps to make a link between expressive motion qualities and emotions which avoids the arbitrary features selection. This paper builds on our preliminary works on the Microsoft Kinect-based gesture recognition based on the LMA method [4,5], where gestures were performed without considering human states or emotions. We only use kinematic features to characterize human motions for a robotic application. Now, we add Effort LMA component which describes the qualitative aspects of the movement by identifying the type of energy used in the movement. Then, we evaluate our descriptor on both motion and emotion recognition through our dataset. Our dataset is composed of 5 gestures (Dancing, Moving, Waving, Pointing, and Stopping) performed with 3 emotional expressions (Happy, Sad, Angry) and Neutral state. Each gesture is repeated 5 times by 11 different subjects. The dataset is acquired from the MVN Awinda system [6]. It's about a 3D motion capture data composed of 17 motions trackers attached to the different body segments. The main contributions of our work are:

- An LMA-devoted motion representation for classifying motion and emotion with the RDF method.

- A publicly accessible motion dataset of 5 expressive motions performed with 4 different emotions selected from the Arousal-Valence model (Happy, Angry, Sad, and Neutral).
- Studying the importance of our motion descriptor to characterize each emotion via a machine learning method.
- Evaluating our recognition system which is based on an automatic learning method against a human perception approach.

The rest of the paper is organized as follows: Section 2 presents the related works in human expressive motions field. Section 3 provides the technical details of the proposed approach. Section 4 presents two main experiments based on machine learning and statistical methods, respectively. The paper ends with conclusion and perspectives for future work.

## 2 Related works

Emotions are the intrinsic moods which are not independent of physiological changes. In line with recent developments in computer vision and multimedia technologies, there has been much interest in emotion understanding. In this regard, the facial expression is one of the popular methods which has been used in various studies [7]. However, the impact of the age was the main drawback of this method. Several studies have been done to reduce the errors that occurred due to the age factor such as [8–10]. Another modality was been considered in the emotions analysis field which is the speech. Kamaruddin et al. [11] have detected emotions through the vocal channel. They studied the correlation between driver behavior states and speech emotions for analyzing driver behavior. In their work, the extraction of the Mel Frequency Cepstral Coefficient accompanied with Multi-Layer Perception and fuzzy neural network classifiers are used. However, this modality is not always available, for example, when affective expressions are estimated from a long distance, or in a noisy environment. So in this regard, few researchers have tried to enhance the accuracy of emotion recognition systems by using affective multiple data [12]. Several studies incorporate facial expression and audio information either as a feature extraction [13] or as a feature selection [14]. In the recent years, more attention has been paid to use the body movement for transmitting emotions. Body movements were represented by a set of activities usually used for 3D human action recognition [15, 16]. However, another field was appearing that used the body movement to convey human emotions. Many authors have confirmed the advantage of this modality in the emotional field, such as, Argyle et al. [17] who have shown that the body movement is as strong as facial expressions in conveying emotions. Gelder et al. [1] have suggested that body expressions allow for successful emotion conveying across long distances. Kleinsmith et al. [18] have compared between bodily and facial expressions, and demonstrated the importance of bodily expressions for non-verbal communication in Human-Computer and Human-Robot interactions. To understand the body ways of expression, Laban [19] formalized movement qualities relying on a body movement analysis system, called LMA. It is particularly useful in describing qualitative aspects of the movement crucial to expression and communication. The LMA categorizes movement descriptions into four main components: Body, Effort, Shape, and Space.

Several researchers have applied this system for emotion analysis through body movement. The LMA was used for several purposes, such as *generating expressive motions* which consists in regulating LMA factors in some gestures to generate an expressive motion. Chi et al. [20] developed an EMOTE system and applied Effort and Shape components to generate expressive gestures for virtual agents. Same as Zhao et al. [21] used the same model to produce natural gestures for embodied communicative agents. They extracted motion qualities inspired by Effort and Shape components. Amaya et al. [22] introduced a model to generate an emotional movement from a neutral movement. They extracted emotional transforms using a signal processing technique relying on two components (the speed and the spatial amplitude transform). These parameters were applied to existing motions and generate two emotions (anger and sadness). Masuda et al. [23] proposed a motion rendering system that adds a target emotion to arbitrary movements. They extracted movement features based on the LMA method. They studied the correlation between Laban's features (Space, Time, Weight, Inclination, Height, and Area) and four emotions (Pleasure, Anger, Sadness, and Relaxation). Duripinar et al. [24], improved the EMOTE model proposed by Chi et al. [20]. They grounded the association between the personality model OCEAN and the body motion and applied this association to animate expressive characters with personality. Shafir et al. [25] asked people familiar with LMA to identify a set of motor characteristics that predict the enhancement of specific emotions like anger, happiness, and sadness. The purpose behind it is to be able to regulate their emotions through their motor behavior. Also, the LMA method was used for *motion retrieval* which consists in indexing and searching complex motions in large motion databases. Kapadia et al. [26] proposed a method for indexing and retrieving motions from a large motion database. To search for complex motions, they extracted motion cues based on LMA components composed of structural, geometric, and dynamic features. Aristidou et al. [27] introduced an approach for content-based retrieval of dance motion capture data inspired from LMA qualities. Also, the LMA model was used for *style transfer* which consists in carrying over the style from one avatar motion onto another. Airstidou et al. [28] used an iterative time warping to map from the input to the output sequence. They transformed a motion sequence into a new motion style while retaining its original content. In another context, for *emotion recognition* the LMA model was applied by several authors, such as Zachartos et al. [29] who extracted motion features based on only the Effort LMA component for emotion recognition in a game scenario. They used the Multi-Layer Perceptron classification algorithm to recognize four emotions (Meditation, Concentrate, Excitement, and Frustration). Recently, the LMA technique was used for characterizing emotions through a specific action, such as waving [30, 31]. The authors asked participants to perform the gesture "waving" with four emotions, angry, happy, sad, and polite. They extracted three Effort qualities, time, weight, and flow and studied the relationship between these qualities and the four emotions when performing the waving gesture. For "walking" task, Cimen et al. [32] quantified LMA components to extract posture, dynamic, and frequency-based descriptors for human motion. They used the SVM method for emotions classification when walking in different emotional states (Angry, Happy, Sad, and Relaxed). These studies mostly focused on emotions expressed in a single movement.

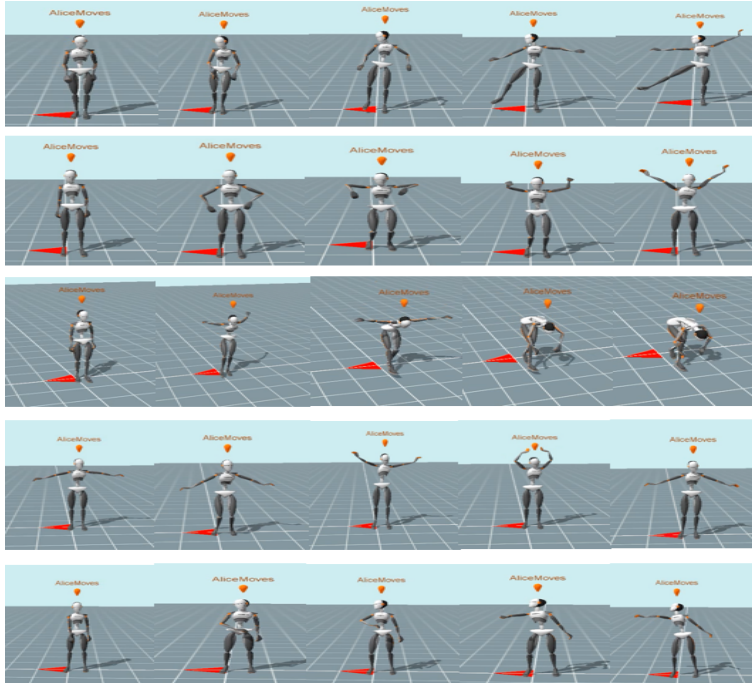
Most of the researchers who have studied expressive gestures have focused on a specific action such as waving, walking, Knocking, etc. and focused on the LMA Effort component as it is the same movement and only the intention of the movement differs. For our case, we have proposed a generic motion descriptor that can describe the quantitative and qualitative aspects of the movement so allows to differentiate different movements and also a same movement performed with different emotions. Moreover, most researchers who have adopted the LMA method worked in the field of music and dance and were forced to ask professional people for the construction of their dataset and sometimes make some training courses [28]. For our case, we chose to create a publicly accessible motion database easy to use by everyone, so we did not need to professional people to construct our dataset or to test an unknown expressive motion. Our work advances the previous researches in human expressive motions recognition. We proposed a descriptor motion to characterize both human motions and emotions. We created a dataset composed of limited gestures (not free) performed with different emotions selected from the Circumplex model of affect (Happy, Sad, Angry, and Neutral). We studied the importance of our motion features in characterizing each emotion. We proposed a human approach based on human perception and ratings to evaluate both emotions and motion descriptor. The obtained results from human classifier were used to evaluate the accuracy and the adequation of our recognition system based on automatic learning method.

### 3 Methods

The whole process of our method consists of four main parts. The first part is data collection using the Xsens capture motion for our emotional dataset. The second part is the feature extraction using the LMA method to convert raw data into a robust motion descriptor. The third part is the classification of expressive motions. Finally, we study the importance of each motion feature to express each emotion. These last two steps are performed using two different methods, the first one is based on a machine learning algorithm (RDF), and the second one is based on a statistical study based on viewers' perceptions.

#### 3.1 Data collection

**Participants** 11 subjects (five males and six females) from the University of Evry Val d'Essonnes ranged in age from 27 to 36 years old ( $M=29.85$  years,  $SD=2.47$ ) participated in this study. They are master and phd students in the information technology field. They were paid for taking part in our research. They expressly allowed us to publish their movements on the internet and to use their motion capture data for research purpose. In order to keep away from having any overstatement feeling emotions, we did not choose professional actors, whom, thanks to theatrical training, know the manner of using their body to demonstrate effect through actions. Our dataset has then the ability to be enriched by everyone without the need for professional actors.



**Fig. 1** Expressive gestures dataset in order from top to bottom: Dancing, Moving, Stopping, Waving and Pointing.

**Expressive gestures** Our emotional dataset is composed of five expressive motions (Dancing, Moving, Waving, Pointing, and Stopping), illustrated in Figure 1. We choose such motions for our future application dedicated to the teleoperation of NAO robot with expressive control motions. For each category of movement, we captured neutral motions, as well as motions performed with emotions: Happy, Angry, and Sad. Unlike other affects such as surprise or disgust which could not be easily identified especially by learning method, the common point amongst these three chosen affects is their ability to express over a relatively long time, a mood or an emotional state which could be recognized easier by the population and also by the learning method. To help the participants to express well their emotions, we play different music corresponding to each situation, and we propose scenarios presenting emotional situations. Each participant was asked to read the proposed scenario, take the time to feel the situation, and perform the action 5 times. We were careful to give the participants only verbal instructions for expressive motions, rather than performing the movements ourselves. Examples of scenarios proposed to participants when performing an expressive gesture (Figure2):

**Happy emotion:**

- You have passed all your courses in the University and the same day that you get your results you have a phone call from a company that interviewed you, and you understood you got also your favorite job.

- You are in a good relationship and the birthday of your partner is close and you are planning to have a party but you do not have enough money, suddenly you receive an email which informs you of winning in a lottery.

**Sad emotion:**

- For a popular festival you booked your ticket and hotel from 6 months ago but because of a friend you miss your flight and you can not arrive at your favorite festival.
- You are in a new country that you have never been there and you use your phone to find your address but suddenly it sticks and stops working, you do not know their language and they can not speak English or your mother tongue at all.

**Angry emotion:**

- You have a job interview which you have been waiting for a long time. Today is 24/03/2018 and you forget that tomorrow, the date of your interview, clocks are turned forward an hour because of Daylight Saving Time. So, you wake up late, nearly an hour, you dress up quickly and get in your car but it does not work. You have no other choice than public transport. Finally, in the subway station, you find that the drivers are on a strike.
- You have your own restaurant but you find out that your business is not making enough money. You decide to discover the problem, so you introduce yourself as a new employee to your employees. On the first week, you notice that the restaurant is opened lately and, as a result, there is always a long delay for customer's orders. On the other hand, you see the unpleasant behavior of waiters with customers which is due to their dissatisfactions with their incomes. In the end, you are very angry about the inefficiencies of the managers who are hired by you.

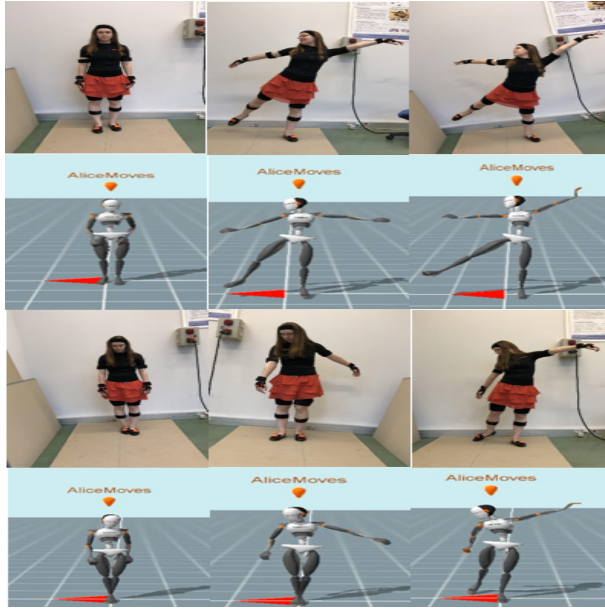
During the recording sessions, the order of scenarios, emotions, and motions was randomized from one participant to another. For each participant, we collected the following files: (1) a .avi video recorded by a camera containing the video of the participant's motion, (2) a .csv file including all motion information produced by the Xsens, and (3) an avatar animation video reproducing the motion of the character. In total, we obtained 1100 motion capture sequences (11 participants  $\times$  5 motions  $\times$  4 emotions  $\times$  5 times).

**Motion capture** For each performed gesture, 3D motion sequence of the person's full body was recorded by the Xsens sensor, as shown in Figure 2. The Xsens estimates the 3D position and orientation of each body joint using 17 inertial trackers.

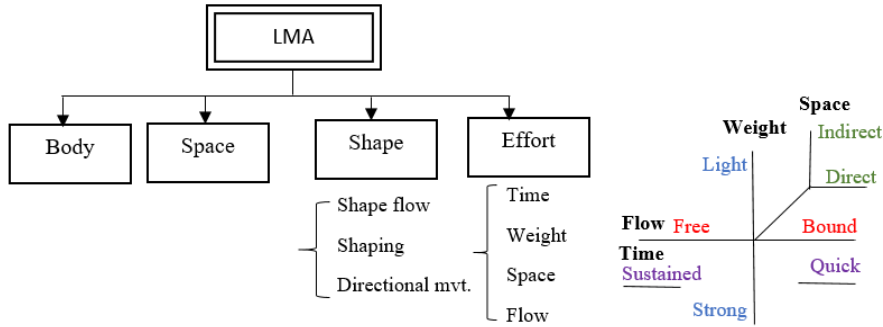
### 3.2 Feature extraction

In the direction of describing human behavior, Laban Movement Analysis (LMA) method is used. The extracted body features are made up of four LMA components (Body, Space, Shape, and Effort) and each of them is described through



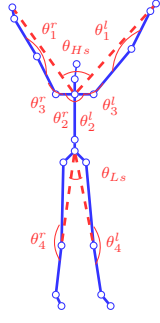


**Fig. 2** Dancing gesture expressed with Happy and Sad emotions reproduced by a virtual avatar.



**Fig. 3** LMA Components (Body, Space, Shape, and Effort).

a specific explanation that quantifies and qualifies various attitudes of dynamic motions (Figure 3). **Body** component has the responsibility of highlighting the body part which is moving, making the connection between the moving parts and taking into consideration the issues of locomotion and kinematics. For this category, we describe the organization and connection between the different joints (see Figure 4). We consider two parts, the upper and lower part. For the first one, the extension of the different joints is described by computing the following angles in the left and right parts respectively: between hands and shoulders ( $\theta_1^l, \theta_1^r$ ), between elbows and hips ( $\theta_2^l, \theta_2^r$ ), between elbows and shoulders in the symmetrical part ( $\theta_3^l, \theta_3^r$ ). We also calculate the distances between two hands ( $d_{Hs}$ ) as well as the distances between the shoulder center and both hands ( $d_{shc, lh}, d_{shc, rh}$ ). For the lower part of the body, the extension of the knees has been described with the an-



**Fig. 4** Body characteristics.

gles between feet and hips ( $\theta_3^l, \theta_3^r$ ). These two characteristics allow to characterize specific actions like crouch or hide gestures. We also characterize the opening of the legs with the angle computed between two knees  $\theta_{L,s}$ . Mean, standard deviation, and range of body features are computed to quantify the Body component.

**Space** component describes the trajectory performed by the participant's body parts during an action. We compute the length ( $L$ ) of the trajectory made by the upper extremities including head, left and right hands joints ( $J_t$  is the 3D joint position captured at frame  $t$ , and  $T$  is the number of frames in a motion sequence.

$$L = \sum_{t=1}^{T-1} \|J_{t+1} - J_t\| \quad (1)$$

**Shape** component can be defined as a collection of the properties that appear from body and space components. This category expresses the changing form of the body that is either motivated by self or the environment. The three factors of Shape are Shape Flow, which represents a relationship of the body to itself; Directional movement, which defines a relationship of the body toward some part of the surrounding environment; and Shaping, which describes the qualitative changes in the shape according to three-dimensional planes. For the *Shape Flow*, we compute the convex hull volume ( $V$ ) of the 3D skeleton based on Quickhull algorithm [33] in order to describe body extension. Mean, standard deviation, and range of the 3D convex hull volume feature are computed to quantify the Shape Flow factor. For *Directional* factor, we focus on the upper extremities including, head, left and right hands. We describe their pathway by computing the gradual angular change  $\phi$  occurring between two successive frames, defined as follow:

$$\phi_{J_t} = \arccos\left(\frac{\overrightarrow{J_{t-1}J_t}}{\|\overrightarrow{J_{t-1}J_t}\|} \cdot \frac{\overrightarrow{J_tJ_{t+1}}}{\|\overrightarrow{J_tJ_{t+1}}\|}\right) \quad (2)$$

This equation describes the local curvature of joints trajectory. From this angle we derived the curvature feature ( $C$ ) defined as:

$$C = \sum_{t=2}^{T-1} \phi_{J_t} \quad (3)$$

This index tends to 0 in straight-Line trajectories cases and changes to a very high value in curved paths. For *Shaping* factor, we describe movements according to

three planes, Horizontal (sideways movements), Sagittal (forward/backward movements), and Frontal (upward/downward movements). We describe body inflation according to the three planes by computing average distances of all skeleton joints  $[J_{t,i}]_{X,Y,Z}$  ( $t = 1, \dots, T$  and  $i = 1, \dots, N$ ) with respect to the spine joint ( $J_{1,s}$ ) at initial frame.

$$D_H = \frac{1}{T} \sum_{t=1}^T \left( \sum_{i=1}^N \text{sqrt}(( [J_{t,i}]_X - [J_{1,s}]_X )^2) \right) \quad (4)$$

$$D_F = \frac{1}{T} \sum_{t=1}^T \left( \sum_{i=1}^N \text{sqrt}(( [J_{t,i}]_Y - [J_{1,s}]_Y )^2) \right) \quad (5)$$

$$D_S = \frac{1}{T} \sum_{t=1}^T \left( \sum_{i=1}^N \text{sqrt}(( [J_{t,i}]_Z - [J_{1,s}]_Z )^2) \right) \quad (6)$$

The intention and dynamic qualities of the movement, the feeling tone, the texture and the manner of consumed energy during movement are explained by the **Effort** component. This component is generally associated with the change in emotion or mood, hence, it is useful for motion expressivity. This factor is required for the description of emotions. We focus on the upper body part since this is the part that moves the most when expressing an emotion, especially the following four joints: head, spine, left and right hands. The Effort component consists of four factors: Space, Weight, Time, and Flow which are explained as follows: *Space* expresses the quality of active attention to the surroundings. It is divided into two categories, Direct (when the action is direct the attention is on a single point in space, focused and specific) and Indirect (giving attention to multiple directions in the space, multi-focused and flexible). We compute the Straightness index ( $S$ ) of upper body joints trajectories ( $Sh$  for head,  $Ss$  for spine,  $Sl$  for left hand, and  $Sr$  for right hand) as the proportion between the Euclidean distance of the straight trajectory between the positions at the first and last frame ( $D$ ), and the sum of distances between successive frames ( $L$ ) defined in the Equation 1.

$$S = \frac{D}{L} \quad (7)$$

In the direct movement, we obtain a straightness index close to 1, and close to 0 in indirect case. *Time* is the interior attitude of the body toward the time, not the period or the length of the movement. Time category can be expressed as sudden (quick, urgent, unexpected, surprising) and sustained or stable (stretching in the time, continues, slowly). So to quantify this factor we compute the mean, the standard deviation, and the range of velocity ( $v$ ) of upper body joints (head ( $vh$ ), spine ( $vs$ ), left and right hands ( $vl, vr$ )). *Weight* involves the strength or power with which a movement is performed. It can be strong or light. Strong movement requires force and acceleration while light movement is characterized by an invariant rhythm of motion, so lower acceleration values. We compute the mean, standard deviation, and the range of the acceleration ( $a$ ) of the head ( $ah$ ), spine ( $as$ ) and hands ( $al, ar$ ) to quantify the Weight factor. *Flow* involves the continuity of the movement. Its categories are bound (controlled movement) and free (unstoppable, liberated). We compute the yaw and pitch ranges of the head

(*range\_yawh, range\_pitchh*), spine (*range\_yaws, range\_pitchs*), left hand (*range\_yawl, range\_pitchl*), and right hand (*range\_yawr, range\_pitchr*) motions. Bound movement can be described as a movement which can be stopped at any moment, whereas free or smooth movement presents a continuity in the motion. Once we have quantified all LMA factors, we obtain a descriptor vector composed of 85 features which will be the input of the classification method presented in the next section.

### 3.3 Expressive motion recognition and analysis using the learning method

#### 3.3.1 Recognition

For expressive motions classification, we used the Random Decision Forest method (RDF) [34], since it has shown its effectiveness in characterizing human actions in our preliminary work [4]. This is also explained by the fact that this method is part of the family 'Ensemble od methods' whose general principle is to build a collection of classifiers, then to aggregate the set of predictions, which increase classifier accuracy as explained by Dietterich [35]. Also, the RDF method uses the randomization in the induction process, with the two methods: bagging [36] and random subspaces [37], in order to build base classifiers different from each other, and thus introducing diversity among them. This makes the model more general and helps to combat the overfitting problem which may have appeared in other models such as SVM and its enhancements [38, 39]. Another great quality of the random forest algorithm is its ability to measure the relative importance of each feature on the prediction. Let the feature vector of training data be  $v = \{f_i\}$ ,  $i = 1, \dots, d$ , where  $d$  is the number of features for each sample. At each node, the best split is chosen using a random sample of  $p$  features from  $d$ . Consider a node  $k$  comprising  $S_k$  samples, split into left and right child nodes with subsamples of  $S_{kl}$  and  $S_{kr}$ , respectively, the tree is then grown by selecting the splitting conditions that maximize the purity of the resulting tree. Gini index  $I(S_k)$  is used to select the feature at each internal node  $k$ . The amount of homogeneity gain achieved by the splitting node  $k$  in feature  $f$  can be evaluated in the following equation:

$$G(f, S_k) = I(S_k) - \sum_{i \in l, r} \left( \frac{|S_{ki}|}{|S_k|} I(S_{ki}) \right) \quad (8)$$

Where  $I(S_k) = 1 - \sum_{j=1}^l \left( \frac{S_k^j}{S_k} \right)^2$ ,  $l$  is the number of classes in node  $k$ ,  $S_k^j$  denotes the number of learning samples which belong to class  $j$  at node  $k$ . Therefore, after several selections of  $f$ , the one producing the lowest value of Gini index is picked as the split criterion for the node. In the testing step, each test sample is simultaneously pushed through all trees, starting from the root, and assigning the data to the right or left child recursively until a leaf node is reached. Finally, the forest chooses the classification having the majority of votes of each decision trees.

#### 3.3.2 Features importance

The features importance step consists in keeping the smallest subset of most relevant features for motion representation to achieve a good compromise between

accuracy and runtime in the classification process. The measure of relevant features is returned by the RDF method. During the training phase, each tree is grown using a different bootstrap sample from the original training data, leaving 1/3 as OOB (Out Of Bag) to estimate the prediction error of OOB. The importance of the feature  $f_i$  is measured as the difference between the OOB error rate of each tree before and after permuting  $f_i$ . Our feature selection method is presented in the Algorithm 1.

---

**Algorithm 1:** Feature selection process.
 

---

**Input** :  $v_0 = \{f_i\}, i = 1, \dots, p$  ▷  $v_0$  is the whole feature set.  
**Output**:  $v^* = \{f_j\}, j = 1, \dots, p^*$  ▷  $v^*$  subset of most relevant features.

- 1  $k = 0$
- 2 **while**  $|v_k| \geq 1$  **do**
- 3     Compute and record OOB error rate:  $E_k(v_k)$
- 4     **for**  $i = 1$  **to**  $p$  **do**
- 5         Compute  $I(f_i)$  ▷ Importance of each feature in  $v_k$ .
- 6     **end**
- 7     Sort  $\{f_i\}$  in descending order according to values of  $I(f_i)$
- 8      $f_{min} = \underset{i}{\operatorname{argmin}}\{I(f_i)\}$
- 9     Apply Tukey's test and select set of features  $\{f_i\}$  that does not lead to a significant change of  $E_k$
- 10     $R = f_{min} \cup \{f_i\}$
- 11     $v_{k+1} = v_k \setminus R$
- 12     $k = k + 1$
- 13 **end**
- 14  $v^* = \underset{l}{\operatorname{argmin}}\{E_l(v_l)\}$  ▷  $v^*$  is the optimal feature subset with minimal OOB error.

---

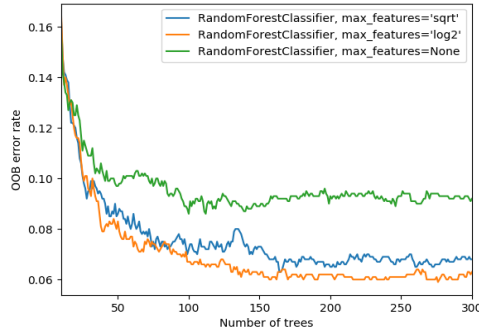
### 3.3.3 Results

**Recognition:** After converting raw data into LMA features we obtained a descriptor vector composed of 85 features. The RDF method was used for motions and emotions classification.

For motions classification we consider all data, 1100 sequences (11 persons  $\times$  5 gestures  $\times$  4 emotions  $\times$  5 repetitions). In this experiment, we evaluated the robustness of our descriptor to classify motions with different intentions. We used the 3 folds cross-validation method, so data is divided into 3 groups. We loop through 3 groups and each time we use one group as the testing set and the remaining groups as the training set. Finally, the performance measure is obtained by the average of the values computed in the loop. As performance measures, we used the accuracy, precision, recall, and the F-score ( $F$ ), where

$$F = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

$\text{precision} = \frac{TP}{TP+FP}$  and  $\text{recall} = \frac{TP}{TP+FN}$ , TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. We adjusted RDF parameters for best classification results. Most important RDF



**Fig. 5** Optimal RDF parameters ( $ntrees$ ,  $maxfeatures$ ).

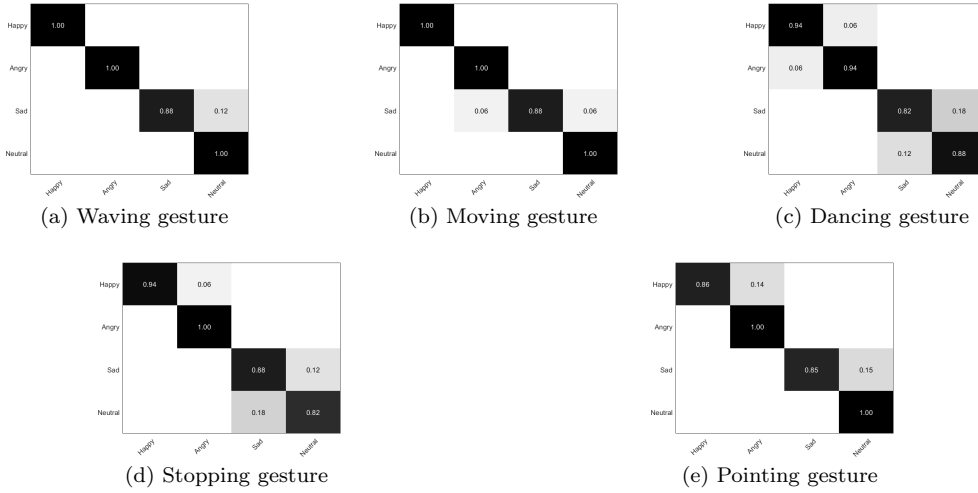
**Table 1** Emotions recognition results (Accuracy, Precision, Recall, and F-score) on the different gestures of our dataset.

| Gestures     | Accuracy (%) | Precision (%) | Recall (%) | F-score |
|--------------|--------------|---------------|------------|---------|
| Waving       | 90           | 90            | 89         | 0.89    |
| Moving       | 78           | 77            | 76         | 0.75    |
| Dancing      | 62           | 78            | 76         | 0.75    |
| Stopping     | 86           | 88            | 86         | 0.86    |
| Pointing     | 95           | 95            | 94         | 0.94    |
| All gestures | 81           | 86            | 81         | 0.80    |

parameters are: the number of trees ( $ntrees$ ) and the number of features to consider when splitting a node ( $maxfeatures$ ). We varied  $ntrees$  starting from 10 to 300 trees, and we tested three values of  $maxfeatures$  (85,  $\log_2(85)$ , and  $\sqrt{85}$ ). The best F-score of 0.83 (+/- 0.07) was achieved when setting  $ntrees = 150$  and  $maxfeatures = \log_2(85)$ . We also confirmed the RDF parameters values by computing the OOB error rate while varying  $ntrees$  and  $maxfeatures$ . As we can see in Figure 5, the two curves of  $maxfeatures = \log_2(85)$ , and  $maxfeatures = \sqrt{85}$  are very close with a very little superiority result of  $maxfeatures = \log_2(85)$ . We identified the minimum value of  $ntrees$  where the OOB error stabilize (around 0,06), we found  $ntrees = 150$ , which confirms the result obtained with the recognition rate measure.

For emotions classification, we divided the dataset into five groups per gesture in order to study emotions recognition for each gesture. Each group is composed of 220 sequences (1 gesture  $\times$  4 emotions  $\times$  11 persons  $\times$  5 repetitions). In each group, we consider 7 subjectif for the training data and 4 subjects for the testing data. Table 1 shows the recognition results obtained in emotions recognition experiment for each gesture (waving, moving, dancing, stopping, and pointing) and for all gestures.

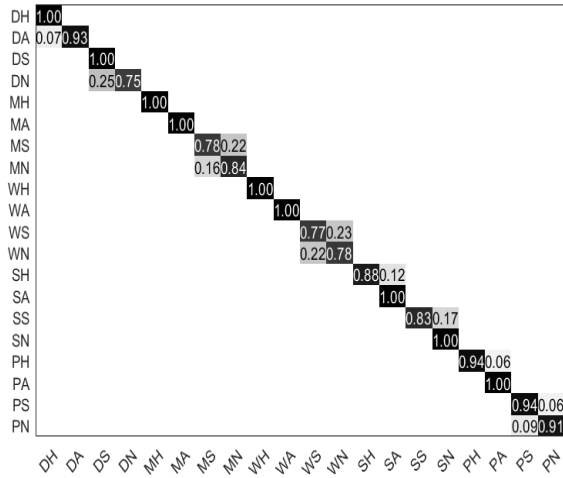
Confusion matrices of the different emotions are presented in Figure 6. The results show that our descriptor vector can distinguish between different motions and even between same motion performed with different emotions. Some expressive motions have been successfully recognized (100%) such as: Waving and Moving gestures with Happy and Angry, and Neutral states, Stopping and Pointing ges-



**Fig. 6** Confusion matrices between expressed emotions (in rows) and perceived emotions (in columns) for each gesture using RDF method.

tures with Angry emotion, and Pointing gesture with a Neutral state. Otherwise, some confusions were obtained in the same motion especially between Happy and Angry emotions, like Dancing gesture, and between Sad and Neutral states in the Stopping gesture. According to these results, we can approve the effectiveness of our motion descriptor in recognizing human motions and emotions through body movement. We also computed the confusion matrix of all expressive motions in Figure 7. So we have 20 classes (5 gestures  $\times$  4 emotions). As we can see, the highest values are marked in the diagonal of the matrix, which confirms the performance of the recognition results. We have no confusion between different gestures, but we can find some confusions in the gesture itself when it was performed with different emotions. Such confusions are obtained in Moving, Waving, and Stopping gestures when feeling Sad and Neutral. Generally, we can say that our descriptor motion has successfully characterized both quantitative and qualitative aspects of motions.

**Features importance:** In the second experiment, we aimed at studying the importance of each motion features to contribute to each emotion. We used the RDF method to select the most relevant features for each emotion based on OOB error rate measures. Here, we are more interested in the qualitative aspect of the movement. Since we consider at each time the same gesture acting with several emotions, so we have the same kinematic features importance. The only difference appears at the level of motion dynamics. For this, we keep only Effort-Shape factors, which describe the different expressive qualities perceived in the motions. Also, to study motion features importance in expressing emotions, we have to consider at each time the neutral emotion as a reference to other emotions (Happy, Angry, and Sad). So, with the RDF method we classified each time one emotion with the Neutral state. Then, we evaluated the importance of each feature in characterizing each emotion by applying our algorithm presented in Algorithm 1 which consists in keeping only the most important features. We obtained the fol-



**Fig. 7** Confusion matrix of all expressive gestures, 5 gestures (D dancing, M moving, W waving, S stopping, and P pointing) performed with 4 states (H happy, A angry, S sad, and N neutral).

lowing results illustrated in Table 2. From this table, we can draw the following conclusions:

The Happy emotion is characterized more by the strength and the urgency of the movement. So, as important features for this emotion, we have the velocity and the acceleration of body articulations in almost all gestures. For example, in Moving gesture with raising both hands, the proposed features importance algorithm gives as results, the velocity and the acceleration of the right and left hands are the most important features in characterizing the Happy emotion (Figure 8).

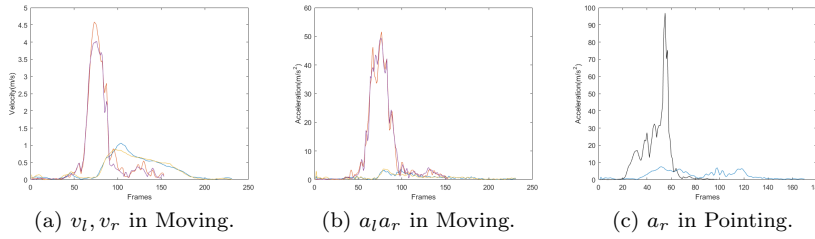
The Angry emotion is more characterized by the strength of the movement in all gestures and the urgency in two gestures (Waving and Moving). Figure 8 illustrates the acceleration feature of the right hand in the Pointing gesture performing with the right hand when feeling Angry comparing to the Neutral state. Another point, according to features importance results, the shape characteristics do not appear significant on the expression of Happy and Angry emotions.

The Sad emotion is characterized by Shape and Effort characteristics in all gestures. For example, in the Moving gesture, we obtained as important features the following Shape characteristics: the convex hull volume of the skeleton, as well as the body inflation according to the sagittal and frontal planes. In the Effort component, urgency (Time), strength (Weight), straightness (Space), and smoothness (Flow) of the movement were considered as important characteristics in expressing the Sad emotion in the Moving gesture. In conclusion, based on the machine learning algorithm we can draw the following relationships between the three emotions and Effort-Shape factors. The **Happy** emotion was described by Time and Weight Effort factors. However no shape factors is important for describing the Happy emotion, we can explain this by the choice of the gestures. In our case, our challenge is to be able to recognize the emotions through limited



**Table 2** Important features results for each emotion across different gestures.

| Expressive motions |          | Motion descriptor          |                   |                                                                                                           |                                                                                                                                  |          |                                                          |
|--------------------|----------|----------------------------|-------------------|-----------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|----------|----------------------------------------------------------|
|                    |          | Shape                      |                   | Time                                                                                                      | Effort                                                                                                                           |          |                                                          |
|                    |          | Shape flow                 | Shaping           |                                                                                                           |                                                                                                                                  | Weight   | Space                                                    |
| Happy              | Moving   |                            |                   | std_vl<br>range_vl<br>std_vr<br>range_vr                                                                  | mean_al<br>std_al,<br>mean_ar<br>std_ar<br>range_ar<br>std_ah<br>range_ah                                                        |          |                                                          |
|                    | Dancing  |                            |                   |                                                                                                           | mean_al<br>std_al<br>std_ar<br>mean_ah<br>std_ah                                                                                 |          |                                                          |
|                    | Waving   |                            |                   | std_vl range_vr                                                                                           | std_al std_ar                                                                                                                    |          |                                                          |
|                    | Stopping |                            |                   | mean_vl<br>std_vl<br>range_vl<br>mean_vr<br>std_vr<br>range_vr<br>std_vh<br>range_vh                      | mean_al<br>std_al<br>range_al<br>mean_ar<br>std_ar<br>range_ar<br>mean_ah<br>std_ah<br>range_ah<br>mean_as<br>std_as<br>range_as |          |                                                          |
|                    | Pointing |                            |                   | range_vl<br>std_vr<br>range_vr<br>std_vs<br>range_vs                                                      | mean_al<br>mean_ar<br>std_ar<br>range_ar<br>mean_ah<br>std_ah<br>range_ah<br>mean_as<br>std_as<br>range_as                       |          |                                                          |
| Angry              | Moving   |                            |                   | range_vl,                                                                                                 | std_al<br>std_ar,range_ar                                                                                                        |          |                                                          |
|                    | Dancing  |                            |                   |                                                                                                           | std_al<br>range_al<br>std_ah                                                                                                     |          |                                                          |
|                    | Waving   |                            |                   | range_vr                                                                                                  | mean_al<br>std_al<br>range_al<br>mean_ar<br>std_ar<br>range_ar<br>mean_ah<br>std_ah<br>mean_as                                   |          |                                                          |
|                    | Stopping |                            |                   |                                                                                                           | std_ar,range_ar                                                                                                                  |          |                                                          |
|                    | Pointing |                            |                   |                                                                                                           | std_ar,range_ar                                                                                                                  |          |                                                          |
| Sad                | Moving   | mean_V<br>std_V<br>range_V | $\frac{D_H}{D_F}$ | mean_vl<br>std_vl<br>mean_vr,std_vr,                                                                      | mean_al<br>mean_ar<br>mean_ah<br>mean_as                                                                                         | S1<br>Sr | range_pitchr<br>range_yawr<br>range_pitchh<br>range_yawh |
|                    | Dancing  | std_V                      |                   | std_vr,range_vr                                                                                           | mean_al<br>std_al<br>range_al<br>std_ar<br>mean_as<br>std_as                                                                     | Sr       | range_pitchl<br>range_yawr<br>range_pitchh               |
|                    | Waving   |                            | $\frac{D_H}{D_S}$ | mean_vl<br>std_vl<br>range_vl<br>mean_vr<br>std_vr<br>range_vr                                            | mean_al<br>std_al<br>mean_ar<br>std_ar                                                                                           |          |                                                          |
|                    | Stopping | std_V<br>range_V,          | $D_F$             | mean_vl<br>std_vl<br>range_vl<br>mean_vr<br>std_vr<br>range_vr<br>mean_vh<br>std_vh<br>std_vs<br>range_vs | mean_al<br>std_al<br>mean_ar<br>std_ar<br>mean_as<br>std_as                                                                      | Ss       | range_pitchr                                             |
|                    | Pointing | mean_V<br>range_V          |                   | mean_vr,<br>std_vr,range_vr                                                                               | mean_ar<br>std_ar<br>range_ar                                                                                                    | Sr       | range_pitchh<br>range_yawh                               |



**Fig. 8** Motion features variation across states (purple/orange curves for left/right hands in Happy emotion; yellow/blue curves for left/ right hands in the Neutral state; black curve for the right hand in Angry emotion).

gestures so for the Happy emotion expressed through a limited gesture we can not have a remarkable difference in the body shape when performing the gesture with happy and neutral states. The only difference can appear in the speed and the strength of the gesture. For the **Angry** emotion, the most important factor is the weight followed by the Time factor. For the **Sad** emotion, Shape and Effort factors are important. Unlike Happy and Angry emotions, shape factors participate well in characterizing the Sad emotion. This can be explained by the fact that performing limited gesture when feeling sad can contribute to a strong contraction in the whole body, so comparing to the Neutral state, we can notice a difference between them. Finally, the Directional movement factor did not have any importance in expressing the three emotions We can say that this factor is not important in such case when performing limited gestures with emotions.

### 3.4 Expressive motion recognition and analysis using the statistical method

#### 3.4.1 Recognition

The purpose of this study is to analyze human emotions based on the human perception. We will see how can a person recognize human emotions through her body movement. We tried to make an external evaluation of the emotions expressed in the videos. We recruited 10 viewers (5 men and 5 women) from the University of Evry Val d'Essonne ranged in age from 28 to 37 years old ( $M=30.9$  years,  $SD=3.16$ ). They watched recorded videos and evaluated each emotion using the 5-item Likert scale (1=strongly disagree, 3=neutral, 5=strongly agree) to determine whether they recognized or not the expressed emotion. For a reliable evaluation, expressive motions in videos were reproduced by a virtual avatar as shown in Figure 1. This help viewers to make scores without being influenced by some factors like facial expressions, gender, etc.

#### 3.4.2 Features importance

Once the evaluation of the emotions was completed, we proceeded to the evaluation of motion features in each observed video in order to study the correlation between emotions and body features. In this part, we focused on the Effort-Shape

components because they are the two major direct specifications or indications of expressive human movements. So, we have 7 factors to evaluate (Shape Flow, Shaping, Directional Movement, Time, Weight, Space, and Flow). We asked another group of 10 viewers from the University of Evry Val d'Essonnes (age from 28 to 31 years) to watch the recorded videos and evaluate each selected features using a 7 point Likert scale as follows:

- Shape Flow: convex hull volume of the body (1=Very shrinking, 7=Very Growing)
- Shaping: distance between spine joint and body extremities according to three planes (1=Very Contracted, 7=Very Extended)
- Directional Movement: the pathway of the body joints (1=Very Curvilinear, 7=Very Rectilinear )
- Time: speed of movement (1=Very Sustained, 7=Very Quick)
- Weight: the strength of movement (1=Very Light, 7=Very Strong)
- Space: straightness of movement (1=Very Indirect, 7=Very Direct)
- Flow: smoothness of movement (1=Very Free, 7=Very Bound)

For example, to rate the Shape flow factor we should look at the development of the volume of the 3D skeleton convex hull. If you perceived that the skeleton volume has been strongly increased throughout movement you rate the Shape Flow factor with 7, and if you perceived that it has been strongly decreased you make 1 as a score.

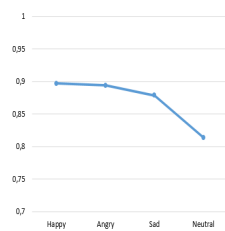
### 3.4.3 Results

#### Inter-viewer reliability

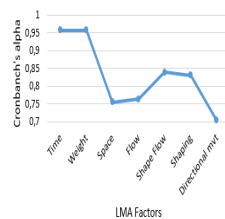
The verification of the consensus among viewers in their decisions is needed in the direction of the emotions perception and the motion features rating with Likert scale. Thus, we used the Cronbach's alpha coefficient to estimate the reliability of the scales given by the viewers when evaluating emotions and motion features. In other words, the Cronbach's alpha is one way of measuring the strength of consistency between viewers decisions. The formula used to compute the Cronbach's alpha coefficient is:

$$\alpha = \frac{n}{n-1} \left( 1 - \frac{\sum_{i=1}^n V_i}{V_T} \right) \quad (9)$$

where  $n$  is equal to the number of items,  $V_i$  refers to the variance of scores on each item,  $V_T$  represents the total variance of overall scores. As we can see in Figure 9, the Cronbach's alpha coefficient is always higher than 0.8 (Happy (0.897), Angry (0.894), Sad (0.879), Neutral (0.814)) which means, according to [40], there is a high consistency in the ratings given by the viewers for the perception of emotions. We also measure the reliability of scales given by the viewers for Effort-Shape factors (Time, Weight, Space, Flow, Shape flow, Shaping, and Directional movement). According to results illustrated in Figure 10, Cronbach's alpha coefficient is higher than 0.7 (acceptable level) for all features rating. So, we can say that viewers were self-consistent when evaluating motion features in our emotional dataset, especially, Time and Weight factors with a reliability coefficient of 0.958. This result shows a strong homogeneity between the viewers when estimating Time



**Fig. 9** Inter-Viewer reliability of the emotions perception using the Cronbach's alpha.



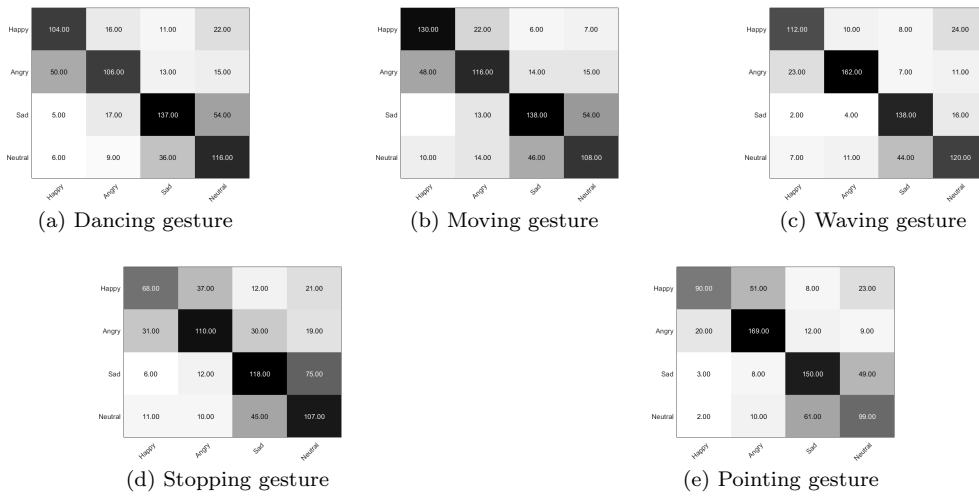
**Fig. 10** Inter-Viewer reliability of Effort-Shape features rating using Cronbach's alpha.

and Weight factors. However, there is a less important but acceptable correlation between the viewers when rating Space, Flow and Directional movement factors with Cronbach's alpha of 0.754, 0.763, and 0.703, respectively. For Shaping and Shape Flow factors, there is a good agreement in the ratings given by the viewers with Cronbach's coefficients of 0.830 and 0.839, respectively.

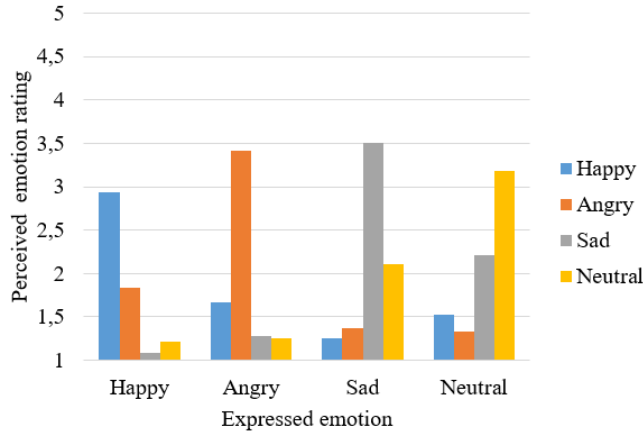
**Recognition:** For emotions recognition based on viewers ratings, we take the 1100 videos for each emotion perception. After finishing the emotions rating by the viewers using the 5-Likert scale, for each emotion perception we consider a recognized emotion when the scale is superior to 3 (neutral). We present the confusion matrix between expressed emotions and perceived emotions in Figure 11. The diagonal cells correspond to the number of correct emotions recognition reported by the viewers (scale > 3). Off-diagonal cells contain frequencies of incorrect classifications. We can conclude that the viewers have correctly classified most emotions in each gesture with some confusions occurred especially between Happy and Angry emotions where they both have high arousal but only differ in the valence, and between Sad and Neutral states. So, based on viewers evaluation we found the same results obtained with the RDF classifier.

After, we considered all expressive gestures and we computed the mean of viewers ratings in each expressed emotion. As we can see in Figure 12, the Happy emotion was successfully recognized by the viewers with some confusions with the Angry emotion and the lowest mean rating was associated with the Sad emotion. For the expressed Angry emotion, the highest mean rating was achieved in the perception of the Angry emotion followed by the perception of the Happy emotion. Sad and Neutral states were well recognized by the viewers in different motions with a bidirectional confusion between the two emotions by certain viewers.

**Features importance:** To study features importance in characterizing human emotions, we measured the correlation between Effort-Shape factors and emotions mean ratings by the viewers using the Pearson correlation coefficient. Pearson



**Fig. 11** Confusion matrices between expressed emotions (in rows) and perceived emotions (in columns) for each gesture based on viewers rating.



**Fig. 12** Mean ratings of emotions perception by 10 viewers.

coefficients are used in statistics to measure how strong a relationship is between two variables and referred to as Pearson's  $r$ , with a value ranging between -1 for a perfectly negative correlation and +1 for a perfect positive correlation, with 0 represents no correlation. The Pearson correlation between the variables  $X$  and  $Y$  is calculated as:

$$r_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Table 3 summarizes Pearson coefficients results obtained when measuring the correlation between the four emotions and Effort-Shape factors ratings: For the Happy emotion, we found a positive correlation with the Shape factors (Shape Flow

**Table 3** Pearson's correlation coefficients  $r$  between Effort-Shape factors and expressed emotions ratings (\*\*. correlation is significant at the 0.001 level).

|        |                 | Happy          | Angry           | Sad             | Neutral         |
|--------|-----------------|----------------|-----------------|-----------------|-----------------|
| Shape  | Shape Flow      | <b>0.541**</b> | 0.132           | <b>-0.524**</b> | -0.316**        |
|        | Shaping         | <b>0.542**</b> | 0.194**         | <b>-0.613**</b> | -0.328**        |
|        | Directional mvt | 0.269**        | <b>0.568**</b>  | <b>-0.505**</b> | -0.397**        |
| Effort | Time            | <b>0.555**</b> | <b>0.622**</b>  | <b>-0.795**</b> | <b>-0.554**</b> |
|        | Weight          | <b>0.543**</b> | <b>0.640**</b>  | <b>-0.780**</b> | <b>-0.594**</b> |
|        | Space           | 0.326**        | <b>0.682**</b>  | <b>-0.566**</b> | -0.487**        |
|        | Flow            | -0.316**       | <b>-0.665**</b> | <b>0.559**</b>  | 0.497**         |

$r = 0.541, p < 0.001$ ; Shaping  $r = 0.542, p < 0.001$ ), and the following Effort factors (Time  $r = 0.555, p < 0.001$ ; Weight  $r = 0.543, p < 0.001$ ). So, the Happy emotion was significantly characterized by growing shape and extended body members. It was also associated with quickness and powerfulness qualities. Masuda et al. [23] have also confirmed in their experiment the same relationship between Happy emotion and Effort qualities.

For the Angry emotion, there is a strong positive correlation with the following three Effort factors (Time  $r = 0.622, p < 0.001$ ; Weight  $r = 0.640, p < 0.001$ ; Space  $r = 0.682, p < 0.001$ ) and a positive moderate correlation with the Directional Movement shape factor  $r = 0.568, p < 0.001$ . The Flow factor was negatively associated with the Angry emotion. So, the Angry emotion was strongly characterized by abrupt, strong, direct, rectilinear, and free movements. Some qualities are consistent with those found by Masuda et al. [23]. The authors have resulted in Angry emotion being correlated with quickness and powerfulness. For Happy and Angry emotions, we found a similar relationship with Effort-Shape factors but with different correlations importance. In Effort qualities rating, the Angry emotion was evaluated as significantly faster, stronger and freer. However, in Shape factors rating Happy emotion was evaluated with a more expanded shape.

The Sad emotion was negatively correlated with all Effort-Shape factors except the Flow quality: Shape factors (Shape Flow  $r = -0.524, p < 0.001$ ; Shaping  $r = -0.613, p < 0.001$ ; Directional Movement  $r = -0.505, p < 0.001$ ), Effort factors (Time  $r = -0.795, p < 0.001$ ; Weight  $r = -0.780, p < 0.001$ ; Space  $r = -0.566, p < 0.001$ ). According to Shape factors rating, the Sad emotion was significantly characterized by shrinking shape, contracted body extremities, and curved path. According to Effort factors rating, the Sad emotion was characterized by light, bound, sustained and indirect movement.

The Neutral emotion was associated with the same Effort-Shape qualities as the Sad emotion but with a lower correlation importance. Masuda et al. [23] found that both emotions (relaxed and sadness) correlated with same Effort qualities (slowness and weakness).

#### 3.4.4 Discussion

The purpose of this study is to characterize and recognize expressed human emotions through body movement first based on machine learning method. Experimental results confirmed the performance of the classifier as well as the proposed descriptor movement. Secondly, we considered the human classifier approach as a

form of baseline. So, we studied the robustness of the RDF method in recognizing human emotions and studying discriminate body features comparing to humans perceptions method. We obtained the following conclusions:

- In emotions recognition study, both RDF and human classifiers have confused between Happy and Angry emotions, and between Sad and Neutral emotions. But, as presented in the confusion matrix when using RDF (Figure6) and human (Figure11), we note that the machine learning method was more accurate than observers. This may be due to the limited number of observers who participated in the rating of the different sequences. Also, it can be explained by the type of gesture chosen. Our approach is very challenging, we try to recognize emotions through limited gestures. For example, through performing the pointing gesture via Sad emotion or Neutral the same movement with a somewhat stable rhythm is performed. So, visually it will not be easy for observers to distinguish between the two types of movement.
- In features importance with the human approach, each observer have evaluated the importance of each LMA factor in characterizing each emotion. With the RDF classifier, the algorithm consists in studying the importance of the features in characterizing emotions comparing to the neutral state. Moreover, we applied the Tukey's test to remove redundant features. As presented in the Table 2, the Sad emotion was characterized by Effort-Shape factors, except the directional movement factor. The same result was obtained with the evaluation of the observers in the Table 3. However, for Happy and Angry emotions, the RDF classifier found that the two most discriminative features in characterizing both emotions are the Time and the Weight factors. These results confirm that the proposed method allows us to characterize the emotions and define the important characteristics while optimizing our motion descriptor by keeping the most robust and significant features.

#### 4 Conclusion and perspectives

The present study set out to investigate the perception of 4 emotions expressed through 5 different motions as well as the importance of the features for each expressed emotion. We exploited the LMA method to represent and analyze the expressive content of movements through several description levels. Two studies are examined: the recognition of emotions through predefined gestures and the selection of most important motion features in characterizing each expressed emotion. Two different methods are proposed, a machine learning method based on the RDF algorithm and a statistical method based on viewers ratings. We used our own dataset composed of 5 expressive gestures performed by 11 subjects captured by the MVN Awinda system. The participants in our dataset are all from the University of Evry Val d'Essonnes. We did not choose professional actors in order to achieve a system accessible to everyone. Our dataset has then the ability to be enriched by everyone without the need for professional actors. The result of this work according to both methods is detailed as follows: For emotions recognition step, different emotions are successfully recognized by two methods, learning and statistical. We also confirm the robustness of our descriptor vector in discriminating both gestures and emotions. For features selection step both methods, RDF, and

viewers rating, demonstrate that Happy and Angry emotions are significantly characterized by Time and Weight qualities. They involve strong and fast movements with higher intensity in Angry emotion. The Sad emotion was characterized with all Effort-Shape factors except the Directional movement factor. Our proposed machine learning method which combines RDF and Tukey's test has successfully kept the most discriminative features which lead to the optimization of our descriptor motion. In the statistical approach, despite recruited people non-expert in the Laban Movement Analysis field in order to express and analyze emotions, they were able to distinguish the Laban features in the movements, which allows our framework to be used even by non-experts. On the other hand, we still have the opportunity to enhance the consistency between the viewer's ratings with some training Laban's theories and techniques. Notwithstanding the above limitations, this study confirms the robustness of our motion descriptor to discriminate expressive motions and give us useful information about the relationship between human states and motion features which allows enhancing interactive applications by regulating body features for each desired emotion.

## 5 Compliance with Ethical Standards

- Funding: This study was funded by the Strategic Research Initiatives project iCODE, University Paris Saclay.
- Conflict of Interest: Author Insaf Ajili declares that she has no conflict of interest. Author Zahra Ramezanpanah declares that she has no conflict of interest. Author Malik Mallem declares that he has no conflict of interest. Author Jean Yves Didier declares that he has no conflict of interest.

**Acknowledgements** We would like to thank the staff of the University of Evry Val d'Essonne for participating in our dataset. As well, we would like to thank Mrs. Alice Jourdan for help with data gathering and tabulation. This work was partially supported by the Strategic Research Initiatives project iCODE accredited by University Paris Saclay.

## References

1. Beatrice de Gelder. Why bodies twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3475–3484, 2009.
2. Harald G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28(6):879–896.
3. Joann Montepare, Elissa Koff, Deborah Zaitchik, and Marilyn Albert. The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior*, 23(2):133–152, 6 1999.
4. Insaf Ajili, Malik Mallem, and Jean-Yves Didier. Robust human action recognition system using laban movement analysis. *Procedia Computer Science*, 112(Supplement C):554 – 563, 2017. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 21st International Conference, KES-20176-8 September 2017, Marseille, France.
5. I. Ajili, M. Mallem, and J. Y. Didier. Gesture recognition for humanoid robot teleoperation. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1115–1120, Aug 2017.
6. Xsens technologies. 2005-2014.



7. T. Kanade, J. F. Cohn, and Yingli Tian. Comprehensive database for facial expression analysis. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 46–53, 2000.
8. Young Ho Kwon and N. da Vitoria Lobo. Age classification from facial images. In *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 762–767, Jun 1994.
9. N. Ramanathan and R. Chellappa. Face verification across age progression. *IEEE Transactions on Image Processing*, 15(11):3349–3361, Nov 2006.
10. A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):621–628, Feb 2004.
11. N. Kamaruddin and A. Wahab. Driver behavior analysis through speech emotion understanding. In *2010 IEEE Intelligent Vehicles Symposium*, pages 238–243, June 2010.
12. M. Pantic and L. J. M. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, Sept 2003.
13. L. S. Chen and T. S. Huang. Emotional expressions in audiovisual human computer interaction. In *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No. 00TH8532)*, volume 1, pages 423–426 vol.1, 2000.
14. L. C. De Silva and Pei Chi Ng. Bimodal emotion recognition. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 332–335, 2000.
15. Ye Liu, Liqiang Nie, Lei Han, Luming Zhang, and David S. Rosenblum. Action2activity: Recognizing complex activities from sensor data. *CoRR*, abs/1611.01872, 2016.
16. Ye Liu, Liqiang Nie, Li Liu, and David S. Rosenblum. From action to activity: Sensor-based activity recognition. *Neurocomputing*, 181:108 – 115, 2016. Big Data Driven Intelligent Transportation Systems.
17. M. Argyle. *Bodily communication*. Methuen Publishing Company, London, 1975.
18. A. Kleinsmith and N. Bianchi-Berthouze. Affective Body Expression Perception and Recognition: A Survey. *Affective Computing, IEEE Transactions on*, 4(1):15–33, January 2013.
19. R. von Laban and L. Ullmann. *The mastery of movement*. Macdonald and Evans, 1980.
20. Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler. The emote model for effort and shape. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '00*, pages 173–182, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.
21. Liwei Zhao and Norman I. Badler. Acquiring and validating motion qualities from live limb gestures. *Graphical Models*, 67(1):1 – 16, 2005.
22. Kenji Amaya, Armin Bruderlin, and Tom Calvert. Emotion from motion. In *Proceedings of the Graphics Interface 1996 Conference, May 22-24, 1996, Toronto, Ontario, Canada*, pages 222–229. Canadian Human-Computer Communications Society, May 1996.
23. M. Masuda and S. Kato. Motion rendering system for emotion expression of human form robots based on laban movement analysis. In *19th International Symposium in Robot and Human Interactive Communication*, pages 324–329, Sept 2010.
24. Funda Durupinar, Mubbasir Kapadia, Susan Deutsch, Michael Neff, and Norman I. Badler. Perform: Perceptual approach for adding ocean personality to human motion using laban movement analysis. *ACM Trans. Graph.*, 36(1), October 2016.
25. Tal Shafir, Rachelle P. Tsachor, and Kathleen B. Welch. Emotion regulation through movement: Unique sets of movement characteristics are associated with and enhance basic emotions. *Frontiers in Psychology*, 6:2030, 2016.
26. Mubbasir Kapadia, I-kao Chiang, Tiju Thomas, Norman I. Badler, and Joseph T. Kider, Jr. Efficient motion retrieval in large motion databases. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, I3D '13*, pages 19–28, New York, NY, USA, 2013. ACM.
27. Andreas Aristidou, Efstathios Stavrakis, and Yiorgos Chrysanthou. *LMA-Based Motion Retrieval for Folk Dance Cultural Heritage*, pages 207–216. Springer International Publishing, Cham, 2014.
28. Andreas Aristidou, Efstathios Stavrakis, Margarita Papaefthimiou, George Papagiannakis, and Yiorgos Chrysanthou. Style-based motion analysis for dance composition. *The Visual Computer*, Nov 2017.

29. Haris Zacharatos, Christos Gatzoulis, Yiorgos Chrysanthou, and Andreas Aristidou. Emotion recognition for exergames using laban movement analysis. In *Proceedings of Motion on Games*, MIG '13, pages 39:61–39:66, New York, NY, USA, 2013. ACM.
30. Tino Lourens, Roos van Berkel, and Emilia Barakova. Communicating emotions and mental states to robots in a real time parallel framework using laban movement analysis. *Robotics and Autonomous Systems*, 58(12):1256 – 1265, 2010. Intelligent Robotics and Neuroscience.
31. Emilia I. Barakova and Tino Lourens. Expressing and interpreting emotional movements in social games with robots. *Personal and Ubiquitous Computing*, 14(5):457–467, Jul 2010.
32. Gokcen Cimen, Hacer Ilhan, Tolga Capin, and Hasmet Gurcay. Classification of human motion based on affective state descriptors. *Computer Animation and Virtual Worlds*, 24(3-4):355–363, 2013.
33. C. Bradford Barber, David P. Dobkin, and Hannu Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.*, 22(4):469–483, December 1996.
34. Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, October 2001.
35. Thomas G. Dietterich. Ensemble methods in machine learning. In *Multiple Classifier Systems*, pages 1–15, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.
36. Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, Aug 1996.
37. Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, Aug 1998.
38. Ye Liu, Luming Zhang, Liqiang Nie, Yuqing Chen, and David S. Rosenblum. Fortune teller: Predicting your career path. In *AAAI*, 2016.
39. Ye Liu, Yu Zheng, Yuxuan Liang, Shuming Liu, and David S. Rosenblum. Urban water quality prediction based on multi-task multi-view learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pages 2576–2582. AAAI Press, 2016.
40. M. Tavakol and R. Dennick. Making sense of cronbach's alpha. *Int J Med Educ*, 2:53–55, 2011.