Vitality Forms Analysis and Automatic Recognition

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

28-09-2021 / 11-10-2021

CITATION

Niewiadomski, Radoslaw; Suresh, Amrita; Sciutti, Alessandra; DI Cesare, Giuseppe (2021): Vitality Forms Analysis and Automatic Recognition. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.16691476.v1>

DOI

[10.36227/techrxiv.16691476.v1](https://doi.org/10.36227/techrxiv.16691476.v1)

Vitality Forms Analysis and Automatic Recognition

Radoslaw Niewiadomski, Amrita Suresh, Alessandra Sciutti, and Giuseppe Di Cesare

Abstract—The form of an action, i.e. the way it is performed, conveys important information about the performer’s attitude. Humans may execute the same action in different ways, e.g., vigorously, gently and rudely. This fundamental aspect of the action has been named *vitality form* by Stern. In this paper we investigate spatiotemporal characteristics of different gestures performed with specific vitality forms and we study whether it is possible to recognize these aspects of action automatically. As the first step, we created a new dataset of 7 gestures performed with a vitality form (gentle and rude) or without a vitality form (neutral, slow and fast). Thousand repetitions were collected from 2 professional actors. Next, we identified 22 features from the motion capture data. According to the results, vitality forms are not merely characterized by a velocity/acceleration modulation but by a combination of different spatiotemporal properties.

Next, we perform automatic classification of vitality forms with F-score of 87.3% and we show that vitality forms recognition is also possible for gesture types which are not present in the training set with the average F-score of 74.2%. These results constitute an important step toward enabling artificial agents to understand human attitudes during the interaction.

Index Terms—Vitality forms, Action style, Nonverbal communication, Motion capture, Recognition

I. INTRODUCTION

Social communication requires the ability to correctly interpret people’s nonverbal behaviors and to timely anticipate others’ actions (e.g., gestures). While interacting with other individuals, we can understand their attitudes as well as their intentions. In addition to the goal (what) and motor intention (why), human actions are also characterized by the form (how). Indeed, the same action can be performed in different ways, for example gently, indifferently, calmly, vigorously and rudely communicating our attitudes toward others. These forms of actions have been named vitality forms (VF) by Daniel Stern [1]. They can be expressed through movement (e.g., goal oriented actions, communicative gestures) as well as voice. The study of vitality forms originally began in the field of psychology and neuroscience [2], [3], but it may contribute to other research domains. For example, endowing artificial agents (being humanoid robots or virtual agents) with the ability to recognize and express vitality forms, can make them more believable and allowing them to interact more naturally with humans. This includes both understanding better the state of their interaction partners and expressing more

clearly their intentions and attitudes (e.g., to show politeness) towards them. Moreover, VF models may have important applications also beyond human-machine interaction. In future, such computational models can be used as a part of training dedicated to, e.g., neurodivergent persons who showed reduced ability to correctly perceive and communicate attitudes through VFs [4], [5].

In order to develop models of vitality forms and to provide artificial agents to perceive them, in this paper we study which spatiotemporal features of goal directed behaviors can reveal the vitality form of an action, in order develop their automatic classification. Thus, first we investigate what are characteristics of different gestures performed with a specific vitality form; second we see whether it is possible to recognize their VF automatically.

In more detail, we focus on hand movements and, by using a motion tracking system, we record the 3D positional data of participants performing different gestures with and without VFs. We collect the data using high precision MoCap system. This allows us to perform the thorough analysis of spatial and temporal properties of a gesture endowing different VFs. We chose to focus on goal directed actions performed gently or rudely. We selected the rude and gentle vitality forms as being able to communicate and understand the related attitudes is particularly important for creation of successful human-artificial agent interaction. For instance, while previous studies show the important role of politeness in such interactions (see e.g., [6], [7], [8]), modelling politeness is often limited to verbal communication only. Being able by an artificial agent to detect and communicate gentle attitude with nonverbal behavior may be used to reduce social distance, increase affinity and trustfulness. On the other hand, being able to detect and simulate rude actions will have important applications in social skills training, simulation and serious games (e.g., [9]). Additionally, these two specific vitality forms have been extensively studied in neuroscience, and thus constitute a representative test-bed to investigate VFs, whose processing is also well characterized at the neurophysiological level [3], [10] (see also Section II). In the last step, we perform machine learning experiments aiming at automatic VFs recognition. We also investigate which cues are the most powerful VF discriminators.

The remaining of the paper is as follows. In Section II, we introduce the recent studies on VF, and in Section III we describe relevant works on internal states recognition from body movements. In Section IV, we explain the data collection procedure. Sections V and VI describe feature extraction, data analysis and classification. Finally, in Section VII, we provide conclusions and we discuss the future works.

R. Niewiadomski was with the Department of Psychology and Cognitive Science, University of Trento, Rovereto, Italy, and CONTACT Unit, Istituto Italiano di Tecnologia, Genoa, Italy, e-mail: r.niewiadomski@unitn.it.

Amrita Suresh was with Inria, Grenoble, France, e-mail: amrita9suresh@gmail.com

A. Sciutti was with CONTACT Unit, Istituto Italiano di Tecnologia, Genoa, Italy, e-mail: alessandra.sciutti@iit.it.

G. Di Cesare was with CONTACT Unit, Istituto Italiano di Tecnologia, Genoa, Italy, e-mail: giuseppe.dicesare@iit.it.

Manuscript received April XX, 20XX; revised August XX, 20XX.

II. VITALITY FORMS

During interactions, important information about others' attitudes is carried out by the form of the action. For example, according to the attitude of the person, a handshake gesture can be performed gently or vigorously. Daniel Stern [1] considered them as a Gestalt created by 5 fundamental dynamic components: movement, temporal contour (time profile), force, space intention/directionality. Vitality form, according to Stern, describes "how" an action is performed, representing an important aspect that an observer may capture observing an action executed by others. Vitality form (how) reflects the internal psychological state of the agent, providing also an appraisal of the affective quality underlying the relation between the agent and the action recipient [1]. It is important to note that vitality forms differ from emotions, and especially from basic emotions, in several respects. Basic emotions are short-lasting responses to relevant events, which involve coordinated changes in several subsystems (cognitive, experiential, physiological, behavioral) [11]. Vitality forms reflect the agents' attitudes and they modulate human behavior in a continuous manner. Importantly, an action performed with a specific VF should not be confused with a specific gesture or utterance that may communicate attitudes. For example, the rudeness might be shown by ignoring the conversation partner (e.g., lack of gaze communication, frequent interruptions), by invading the personal space of the interaction partner, or by pronouncing verbal offences, while gentleness might be shown with, e.g., gifts, or, in speech, by using some conventional sentences considered to be polite in specific cultures. These are examples of actions that may communicate the same attitudes, but they do not necessarily use vitality forms (even if they exploit the same modalities). The ability to express and recognize vitality forms allow people to be socially connected. Indeed, during interpersonal relations, the vitality forms contribute to a social communication by modulating gestures or words, while the perception of vitality forms allows the observer to understand the attitude of others and prepare an adequate motor response.

The VFs are loosely related to the concept of interpersonal stances defined by Scherer as "characteristic of an affective style that spontaneously develops or is strategically employed in the interaction with a person" [11]. The VFs are also related to "expressive quality of the movement" [12] and "gesture expressivity" [13], [14]. It is important to note that, differently from the concepts mentioned above, VFs can be communicated with more than one modality: through the observation of a gesture, through voice intonation and even through touch [15], [16].

Extensive research has provided strong evidence of the neural bases of the generation and comprehension of vitality forms in humans. Recent results obtained in several fMRI studies showed that the observation of actions performed with rude and gentle VFs, as well as listening action verbs or interjections pronounced in gentle or rude way, produced the activation of a small part of the brain named dorso-central insula [3], [17], [16]. Interestingly the same area is active also during the expression of vitality forms, e.g., during the imagining and the execution of actions (or the pronunciation

of words) conveying gentle and rude VFs [10], [18]. These findings suggest the existence of a mirror mechanism for action vitality forms and individuates a central role for the dorso-central insula.

It is important to note that also the observation of humanoid robots performing gestures characterized by human-inspired VFs can evoke similar neural responses in the dorso-central area of the observer [19]. This phenomenon has also behavioral consequences. In a behavioral study aiming to assess the role of vitality forms during a human-robot interaction, Vannucci and colleagues [20] investigated whether and how a robotic action and voice endowed with vitality forms may affect humans. Participants were required to take a ball after the robot action (passing the ball) or its verbal request (take the ball). The two requests were performed gently and rudely by the robot. Results showed that both the voice and action expressed by the robot with vitality forms influenced the kinematic parameters (velocity) of the subsequent actions performed by participants. All these findings highlight that in addition to the goal (i.e., *what*) (e.g., moving an object), vitality forms constitute an important aspect of an action (the form; *how*) that is fundamental to interpersonal relations, both among humans and between humans and artificial agents.

III. RELATED WORKS

While to the best of our knowledge there is no works dedicated to the automatic classification of vitality forms, related topics, such as expressive movement qualities recognition are of large interest (see [14], [13] for recent reviews). Several algorithms classify movement qualities according to Laban Movement Analysis (LBA), a notation proposed by a choreographer Rudolf Laban to describe and interpret human movement. The Effort component [21] of LBA describes *how* the body performs movements and it includes four factors: Space, Weight, Time and Flow. Ran and colleagues [22] applied supervised machine learning to detect Laban qualities from Kinect data. They proposed a large set of descriptors composed of 100 features related to Laban's qualities and other 6000 describing the skeleton data. Multitask learning was applied to 18 qualities resulting in F-score of 0.6. Swaminathan and colleagues [23] used a Bayesian fusion approach for identifying the Laban's Shape component from motion capture data. Samadani and colleagues [24] proposed a set of continuous measures of Laban Effort and Shape components. The values of four components: Weight, Time, Space, and Flow are computed from a set of low level features such as position, kinetic energy, velocity, acceleration, and jerk extracted from the motion capture data of hand and arm movements. The approach was validated by measuring the correlation between the algorithm values and the expert annotations. The results were up to 81% on Effort components. Other works proposed computational approaches for other expressive movement qualities such as Fluidity [25], Impulsivity [26], Smoothness [27], Lightness and Fragility [28].

Other related topic is emotion recognition (see recent surveys [29], [30]). Emotions can be recognized from movement data (e.g., gestures) because specific emotions can be

associated to certain combinations of the kinematic properties of the movement (for example, a person may perform more quick and abrupt gestures when she is angry than when she is sad). So, we believe that it can be possible to recognize VFs, by adapting research methodology, used for emotion classification. In this line, Castellano and colleagues [31] classify acted emotions using the low-level movement features (e.g., motion quantity or velocity) extracted from videos. A set of temporal aggregators was applied to these low-level features, which were later classified in terms of four emotions. Piana and colleagues [32] used 3D motion data of full-body movements and defined a number of low-level (e.g., kinematics of a single joint) and high-level (e.g., contraction index, impulsiveness) features, which were learned by an SVM classifier. More recent approaches include the automatic learning of spatio-temporal movement features. For example in [33], the authors classify emotions from full-body movements by using two Convolutional Neural Networks processing in parallel the RGB images obtained from different time intervals of 3D-positional data. Combining different temporal scales into a single architecture improves the classification results. In [34], a 3-layered RNN was used to perform the classification of emotions from motion capture data of daily activities: clapping, drinking, throwing, and waving associated with four emotions: happy, angry, sad, and neutral.

This work is presenting the first ever attempt to classify VFs. The existing methodology (e.g., applied in the past for emotion recognition from the full-body movements), will be adapted to classify rude, gentle, fast, slow, and neutral gestures. This methodology consists of processing the data intervals by extracting the meaningful and easily interpretable features designed by human experts and by applying standard machine learning techniques. Using such hand-crafted features we gain the opportunity to understand the basic properties of gentle and rude movements.

IV. DATA COLLECTION

While several datasets exist for expressive movement analysis (e.g., [35], [28], [36]) to the best of the authors' knowledge there is no large dataset to study VF. The main goal of this data collection is to extract the kinematic properties of seven different actions performed with two VFs: gentle and rude, as well as the same gestures performed slowly, fast and with neutral velocity (baseline condition) ¹.

A previous study conducted on a very small action sample [37] suggested that gentle actions can be distinguished from rude actions because of their velocity. However, not all slow gestures can be considered gentle, and not all fast gestures are perceived as rude. To show that applying VFs is not equivalent to gesture velocity modulation, in this study we introduce two other classes of gestures: *fast* and *slow*. Fast/slow gestures are performed without any particular VF and are characterized just by velocity modulation. We hypothesize (H1) that considering the velocity is not sufficient to distinguish between fast and

rude, nor between slow and gentle actions. Rather, the combination of several features is needed for this task.

Additionally, we introduce the fifth class that we called *neutral* or baseline condition. The gestures performed in baseline condition (at the beginning of the data collection section) have purely functional character (e.g., grabbing an object in a default manner) without any additional velocity or VF modulation. This class is used mainly to control whether the actions intended to be slow/fast are indeed significantly slower/faster than the baseline (control analysis).

Fig. 1. Markers and rigid bodies positions: C1 - wrist, C2 - elbow, C3 - shoulder, C4 - chest.

A. Protocol

Two theater actors (1 male, 1 female) were asked to perform with their dominant (right) hand four gestures that involve additional objects (to grasp; to offer; to drop a game card; to indicate a point of the surface) and three gestures with no additional objects involved (to rise a thumb up; to rise a finger to the mouth in a silence gesture; to point towards the other person) towards another person sitting in front of them (Fig. 2). Additionally, in the case of “grasping” actions and “offering” actions two different objects were used to create more generic action models. In the remaining of the paper we call *gesture types* these seven different actions (in order to distinguish them from vitality forms). All the gestures are performed in a social setting: in front of the actor there is another person sitting, who does not perform any specific action or gesture ².

Each session was organized in 3 steps. For each gesture, the hand's starting and ending positions were pre-defined (see Fig. 2). More specifically, the same starting position (right hand on a thigh) was used for all actions.

In the first step, the 3D positional data of actions performed without any vitality form were recorded, i.e., actors performed each action in a “neutral” way without receiving any instruction regarding the VF or velocity (baseline condition).

In the second step, they were asked to perform the same actions slowly and quickly. In the steps 1 and 2, each action was repeated at least 10 times.

In the third step, actors were presented with four video-clips showing two individuals performing two transitive (pass

¹Neutral - means focusing on the functional aspect of the movement (e.g., passing an object) without adding any additional affective or other communicative intention, neither modulating its velocity or trajectory in other way.

²A video showing some examples of gestures is included as Supplementary Material

Fig. 2. Video frames showing actions performed by the male actor toward the experimenter sitting in front of him. On each image, the gesture starting position is signed with 1, and ending position with 2.

a bottle, offer a packet of crackers) and two intransitive actions (stop gesture, caressing). These video-clips were used in a previous fMRI experiment showing the activation of the brain areas involved in the processing of gentle and rude VFs [3]. Intentionally the videos were presented without pronouncing the labels *gentle* and *rude*. This procedure was chosen to allow the actors to pay attention to vitality forms without being influenced by key-words. It allowed us to control the experimental procedure by excluding the potential variability in interpretation of linguistic terms related to VFs. When watching videos, actors were asked to focus on the attitude of individuals (their internal state) and not just on the kinematic features, with the intent to reproduce their attitudes when performing the other gestures. Next, each actor was asked to perform the actions in presence of a person sitting in front of him/her, expressing the two attitudes observed in the videos. Specifically, each actor performed actions gently or rudely (vitality forms condition). To augment the movement variability in the dataset, actors were additionally asked to perform these actions in different three directions: not only towards a sitting person (0 degrees) (see Figure 3), but also directed to the right side (45 degrees condition), and the left side of the table (-45 degrees condition). Thus, in the step 3, four actions (all apart grasping, offering and silence gesture) were repeated at least 20 times ($10 \times 0\text{ degrees}$, $5 \times 45\text{ degrees}$ and $5 \times -45\text{ degrees}$). The silence gesture was repeated at least 10 times. The grasping and offering were repeated min. 30 times.

Considering the two actors, in total 1051 actions were recorded and 30 actions (2.9%) were discarded from the dataset due to technical problems. The final number of trials per class and gesture type is presented in Table I.

B. Technical Setup

The multimodal data were recorded using the NDI Optotrack Certus system, and 2 webcams. The motion capture

TABLE I
THE NUMBER OF ACTION REPETITIONS PER CLASS AND GESTURE TYPE.

Type	Baseline	Slow	Fast	Gentle	Rude	Total
Grasp	21	21	22	61	60	185
Offer	20	21	21	62	61	185
Drop	21	20	20	41	38	140
Point - Surface	20	20	20	37	39	136
Thumb Up	20	20	20	40	38	138
Silence	19	20	20	20	19	98
Point - You	21	21	18	40	39	139
Total	142	143	141	301	294	1021

(3D positional data, at 100fps) and video (1920×1080 at 30fps) were synchronized using YARP platform [38]. The synchronized playback (see Fig. 3) was realized with the freely available EyesWeb platform [39]. 16 markers were regrouped into four rigid bodies placed close to wrist, elbow, shoulder, and on the torso. Two additional markers were used to refer to the table placed in front of the actor (see Fig. 2 and 3).

V. DATA ANALYSIS

The aim of this analysis was to investigate spatiotemporal cues that characterize the vitality forms. Data pre-processing, feature extraction and statistical analyses were carried out using Matlab software.

Fig. 3. Synchronized (two webcams: frontal and lateral) view generated with EyesWeb software.

A. Data Pre-processing

To reduce the dimensionality of the input data, we computed the centroids of the marker positions on each of 4 locations on the body (see Figure 1). We thus obtained four such centroid positions. We focus on kinematic features of the wrist (C1), and positional properties of the elbow (C2), shoulder (C3) and chest (C4). All actions started from the same initial position and terminated at approximately the same position (depending on the action type). We divided each action into two major phases: Go (from the start to the end position) and Return (from the end position to the rest/start position). Automatic segmentation of the actions was performed by using pre-selected velocity and acceleration thresholds. For statistical analysis, we only consider the Go phase because the actors were instructed to focus on the goal of the action (GO-phase).

B. Feature extraction

We propose 22 features to assess differences among five different action classes (neutral, slow, fast, gentle, rude). Some refer to specific moments of the gesture (e.g., the maximum velocity), whilst others describe the whole gesture (e.g., the suddenness). The features are based on the existing literature on emotion, movement qualities and social relations detection (e.g., [40], [26], [27], [32]).

First of all, we computed standard kinematics features. **Velocity** is taken to be the instantaneous tangential velocity, i.e. the first derivative of the position. Similarly **Acceleration** and **Jerk** refer to the instantaneous values of the first and second derivative of velocity respectively. In all three cases a maximum value is computed over the whole gesture phase, and it is referred as **Velocity Max (VM)**, **Acceleration Max (AM)**, and **Jerk Max (JM)**. These three features are probably the most often used for gesture and movement analysis and classification [41], and previous works showed that the velocity can be relevant also for the distinction between rude and gentle [37]. Additionally the **VM** feature was used to verify the data collection design (i.e., control analysis) and more specifically to check whether gestures that were intended to be slow, were indeed slower than the baseline gestures and fast gestures were actually quicker than the baseline.

We also estimated a gesture **Suddenness (S)** by adapting the method described in [26]. In our case, the value of the Suddenness is a value of the α parameter of the α -stable distribution of the wrist velocity computed on the whole GO-phase. This feature might be relevant, as the rude gestures are expected to be more sudden than other gestures. It is important to notice that while **VM** refers to one instant of the movement, the Suddenness takes into account the velocity profile during the whole gesture. The Suddenness also corresponds roughly to Quick movement of Time component (Quick Vs. Sustained) in Laban's Effort.

The other group of features focus on the trajectory shape. **MPO Wrist (MPOW)** was defined as the maximum deviation of the trajectory of the wrist from a straight line joining the starting and ending points. This can be taken as a measure of how curved the trajectory is. Similarly, **MPO Shoulder (MPOS)** is the maximum deviation of the trajectory of the

shoulder's centroid from a straight line. **Percent MPO Wrist (%MPOW)** is the percentage of the gesture duration from its start to the maximum path offset (deviation) of the wrist. Similarly, **Percent MPO Shoulder (%SPOS)** is the percentage of the gesture duration from its start to the maximum path offset (deviation) of the shoulder. We proposed these 4 features, taking into account various measures of deviation of the movement from the straight line that were used in the past e.g., for emotion classification algorithms [32]. At the same time, a straight movements correspond to Direct movements of Space component in Laban's Effort.

Arc length (AL) is the total length of the trajectory from the beginning to the end/goal of the action. It is obtained by calculating the cumulative sum of distances between points on the trajectory. In our analyses, we considered only the arc length of the Go phase.

We also compute maximum of **Wrist Curvature (CWM)** as well as average (**SMA**), standard deviation (**SMS**) and maximum (**SMM**) of **Wrist Smoothness** according to the procedure described in [27]. All of them measure how smooth the trajectory of the movement is.

We estimated the movement variability [42] with the maximum (**Variability Wrist Max (VWM)**), average (**Variability Wrist Average (VWA)**), and standard deviation (**Variability Wrist Std (VWS)**) of the difference between the actual wrist velocity and the low-pass filtered velocity of the same joint. We expected that the gentle movements would show more smoothness and less variability than the rude ones.

The last group of features is based on the body joints' angles. The angle representation was often used in the past for classification of affective behavior and social signals (e.g., [43], [40]), and thus we expected the features based on angles measurements to be relevant also for VFs analysis and classification. In detail, **Angle Arm-Chest Max (AACM)** is computed as the maximum angle formed by the Elbow-Shoulder-Chest (ESC), in the plane of the ESC, during the whole GO-phase. Additionally, we also compute the average (**AACA**) and standard deviation (**AACS**) of this angle during the whole GO-phase. Similarly, **Angle Elbow Max (AEM)** is the maximum angle formed by the Wrist-Elbow-Shoulder (WES), in the plane of the WES, during the GO-phase. We also compute the average (**AEA**) and standard deviation (**AES**) of this angle.

C. Statistical Analysis

Twenty two features were identified. To see whether they can be used to differentiate vitality forms, mean features values were modelled using a Repeated Measured GLM (general linear model). For each feature, outlier values were calculated and were discarded from the subsequent statistical analysis (> 2.5 SD of the mean). Sphericity of data was verified before performing statistical analysis (Mauchly's test, $p > 0.05$) and the Greenhouse-Geisser correction was applied in case of sphericity violation ($p < 0.05$). All variables were normally distributed (Kolmogorov-Smirnov Test, $p > 0.05$). In total 22 Repeated Measure GLM were carried out, one for each feature. The significance level was fixed at $p = 0.05$. Each GLM

TABLE II

TABLE REPORTS VALUES BASED ON POST HOC COMPARISONS. GREENHOUSE-GEISSER CORRECTION APPLIED. SIGNIFICANT VALUES $p < 0.5$. IN BOLD THE FEATURES FOR WHICH THE RESULTS ARE SIGNIFICANT. THE LETTERS A-E INDICATE SIGNIFICANT DIFFERENCES IN PAIRWISE COMPARISONS.

Name	Baseline (a)	Slow (b)	Fast (c)	Gentle (d)	Rude (e)	F	p
VM	1 bcde	0.5 ace	2.1 abde	0.7 ace	3 abcd	73.4	$p < 0.0001$
AM	7 ce	2.5 ce	22.8 abde	3.6 ce	49 abcd	128.4	$p < 0.0001$
JM	128.8 ce	47.8 ce	788.2 abde	81.4 ce	1737.1 abcd	45.3	$p < 0.0001$
S	1.3	1.7 e	1.6 e	1.8 e	1.1 bcd	5.1	$p < 0.01$
%MPOS	35.1	44.7	37.3	39.7	38.7	2	$p > 0.05$
MPOS	0.1	0.1	0.1	0.1	0.1	1.9	$p > 0.05$
%MPOW	33.1	37.7	39.3	35.5	32.5	1	$p > 0.05$
MPOW	0.08	0.09	0.15	0.55	0.16	1.1	$p > 0.05$
AL	0.46 e	0.47 e	0.4 e	0.49 e	0.6 abcd	14.6	$p < 0.0001$
CWM	28.5 e	7.4 e	65.9 e	12.5 e	268.2 abcd	4.5	$p < 0.01$
SMM	1.7×10^{-7} bd	7.8×10^{-7} acde	0.2×10^{-7} bd	4.2×10^{-7} abce	0.07×10^{-7} bd	35.8	$p < 0.0001$
SMA	1.7×10^{-7} bd	7.8×10^{-7} acde	0.2×10^{-7} bd	4.2×10^{-7} abce	0.07×10^{-7} bd	36.2	$p < 0.0001$
SMS	2.8×10^{-15} bd	20.1×10^{-15} acde	0.2×10^{-15} bd	10×10^{-15} abce	0.07×10^{-15} bd	50.6	$p < 0.0001$
VWM	8.1 ce	7.4 e	14.4 ae	7.8 e	35 abcd	33.5	$p < 0.0001$
VWA	0.12 ce	0.02 ce	0.8 abd	0.05 ce	1.1 abcd	10	$p < 0.0001$
VWS	1 ce	0.5 ce	4.1 abde	0.6 ce	9.9 abcd	53.3	$p < 0.0001$
AACM	80.7 de	79.9 de	82.9 e	84.6 ab	86.8 abc	9.9	$p < 0.0001$
AACA	62.6 e	62.5 e	62.7 e	63.3	63.8 abc	5.4	$p < 0.01$
AACS	16.2 de	15.9 de	17.6 e	19.1 ab	20.6 abc	9.2	$p < 0.0001$
AEM	80.8 d	81.5	80.7 d	82.4 ac	81.6	4.2	$p < 0.01$
AEA	71.5	70.8	72.2	72.2	72.4	2	$p > 0.05$
AES	13.2 ce	12.9 ce	11 abd	13.5 ce	11.4 abd	5.5	$p < 0.01$

comprised the mean value of each feature obtained during the execution of seven actions performed by the two actors (male and female) in different ways (baseline, gentle, rude, slow, fast). Results of the GLM analyses and post hoc comparison are both reported in the Table II.

Regarding the control analysis, using 2 features (peak of velocity and acceleration) one can distinguish between neutral, slow and fast gestures. The gestures labelled *fast* were significantly faster than the gestures labelled *neutral*, which were significantly faster than the gestures labelled *slow*. This means that the actors correctly modulated the gesture velocity.

D. Discussion

Our results show that it is possible to differentiate VF. First of all, gentle and rude actions differ because of the kinematics features (VM, AM, JM). In particular, as expected, the *rude* gestures can be easily distinguished from *gentle* because of the peak velocity. This confirm previous results [37].

Regarding the H1, we have seen that as a set of features is needed to distinguish rude actions from fast actions, and gentle actions from slow actions. In more detail, the 3 kinematics features: peak of velocity (VM), acceleration (AM) and jerk (JM) are higher for rude gestures than for fast gestures, while the overall measure of Suddenness (S) shows that *rude* gestures are more sudden than other gestures [26]. The *rude* gestures were also significantly longer (AL) than any other gestures, including *fast* ones. The significant differences between *rude* and *fast* gestures were also observed for the variability features (VWM, VWA VWS). At the same, time, rude and fast movements were characterized by higher variability (VWA; VWM) than other gestures. Interestingly the two VFs-based gestures *rude* and *gentle* are longer in average than remaining gestures (i.e., fast, slow, and baseline). This shows that VFs influence the trajectory of the movement.

The slow gestures differ from gentle ones, mainly because of features based on arm-chest angle (AACM and AACS), so the features that focus on position of an arm (and not wrist). Additionally the differences between *slow* and *gentle* were observed for smoothness-based features (SMM, SMA). At the same time, gentle and slow gestures were more smooth (SMM, SMA) than other gestures. None of the standard kinematics features (VM, AM, JM) is able to distinguish *slow* from *gentle* motions in our dataset.

Our findings indicate that fast gestures are significantly different from rude actions, and slow actions from gentle actions. At the same time, it is not enough to modify the peak velocity of a gesture to make it *rude* and *gentle*. This confirms recent fMRI results showing that the mere variation of a robot arm peak speed is not sufficient to modulate the activation of the brain area responsible for VF processing in the action observer. Conversely, a more precise re-targeting of human kinematics, considering also more subtle properties of the movement, succeeded in evoking similar activation as the observation of humans expressing VFs in their gestures [19].

VI. VITALITY FORMS CLASSIFICATION

In this section, we provide the classification results of the vitality forms. We use standard classifiers: a) SVM with RBF (SVM-RBF), b) and polynomial kernel (SVM-PL), c) k-NN, d) Multi-layer Perception (MLP) and Random Forest (RF) that were widely used in the past to recognize human internal states from the kinematics data of upper- and full- body nonverbal behaviors (see e.g., [32], [44], [45], [46]).

To train the classifiers, we used 17 features for which the effect of VF was observed in previous Section (see Table II). Five classes were considered: neutral, slow, gentle, rude and fast. Before training, normalization was applied to all the data. All 1021 gestures were used for the classification. The classes are not balanced (see Table I).

TABLE III

ACCURACY, AVERAGE F-SCORE, AVERAGE PRECISION AND AVERAGE RECALL FOR LEAVE-ONE OUT VALIDATION METHOD AND 5 CLASSES.

Algorithm	Accuracy	F-score	Precision	Recall
SVM-RBF	87.4	87.3	87.4	87.5
SVM-PL	85.4	85.4	85.5	85.4
k-NN	86.3	86.3	86.5	86.3
MLP	81.7	81.6	81.6	81.7
RF	84.9	84.8	84.9	84.9

TABLE IV

CONFUSION MATRIX FOR SVM-RBF.

	Base	Fast	Slow	Gentle	Rude
Base	123	3	0	13	3
Fast	1	127	0	1	14
Slow	4	0	104	24	9
Gentle	15	0	23	257	6
Rude	2	9	0	1	282

A. Leave-one-out validation

In the first series of experiments, the leave-one-out method was used. Grid research was performed to tune SVM using consecutive powers of 2 in the range $\{-1, 9\}$. The best performance was obtained for $C = 8$ and $\gamma = 16$. In the case of SVM-PL the best results were obtained for $C = 8$ and $\gamma = 8$. All the results are given in the Table III in terms of average accuracy, F-score, Recall and Precision. In all cases the accuracy is highly above the chance level (29%). Most of the classifiers performed similarly on the dataset, with SVM-RBF obtaining the best results and MLP the worst.

Table IV shows the confusion matrix for the best algorithm (SVM-RBF). Although good results in general were obtained, still some confusions can be observed. It can be seen that gentle movements are quite often confused with the slow ones, but rude gestures are only very rarely confused with fast gestures.

B. Leave-one-gesture-type-out validation

Additionally we ran a series of experiments using the leave-one-gesture-type-out policy. Each time we removed one gesture type (e.g., all occurrences of *grasping* or all occurrences of *thumb up* gesture) from the training set and we used the data of that gesture as a test set. The procedure was repeated seven times, and each time dimensions of the training and testing set were different (preserving the total number of 1021 actions), as the number of trials for each gesture type was slightly different. This procedure simulates the situation in which the previously trained model is used to recognize VF of a new gesture (which is not present in the training set). In each experiment, independent parameters tuning was performed for SVM-PL, and SVM-RBF.

The average results for seven iterations are provided in the Table V. Unsurprisingly the overall results in this case are lower than the results reported in Table III. Still they are highly above the chance level showing that it is possible to recognize the VF even from the gestures which are not present in the training set. When analysing single gesture types, the best results were obtained for *pointing toward a person* and

TABLE V

AVERAGE ACCURACY AND ITS STD. DEVIATION, AVERAGE F-SCORE AND ITS STD. DEVIATION, AVERAGE PRECISION AND RECALL FOR LEAVE-ONE-GESTURE-TYPE-OUT VALIDATION AND 5 CLASSES.

Algorithm	Accuracy	F-score	Precision	Recall
SVM-RBF	74.4 (5.84)	73.5 (5.76)	76.9 (5.28)	74.4 (5.85)
SVM-PL	74.7 (6.92)	74.2 (6.27)	78.3 (5.37)	74.6 (6.94)
K-NN	63.6 (6.78)	63.2 (7.08)	66.7 (6.22)	63.6 (6.82)
MLP	68.0 (6.13)	67.0 (6.05)	71.4 (5.78)	68.0 (6.15)
RF	72.8 (6.89)	71.5 (7.04)	75.3 (5.80)	72.8 (6.91)

the worst for *silence* gesture (see Figure 4). The latter is quite different from any other gesture in the dataset and may require from the actor a specific kinematic modulation to preserve its meaning. Regarding the classification method, both SVMs obtained high results in terms of Accuracy and F-score, while k-NN (that was the second best for the leave-one-out-policy) achieved this time the worst score.

Fig. 4. The F-scores for each type of gesture when using leave-one-gesture-type-out method and three best algorithms.

C. The best features

To compare the relative relevance of different features in assessing the VFs, we performed a set of experiments with only one feature at once. The Random Forest was used, for this purpose, as it does not require time-consuming parameter tuning. The best performance was obtained for standard deviation of smoothness (SMS, Acc=59.1, F-score=56.1), which was followed by maximum acceleration (AM, Acc=54.4, F-score=54.4), maximum smoothness (SMM, Acc=53.8, F-score=53.8) and average smoothness (SMA, Acc=53.8, F-score=53.8), standard deviation of variability (VWS, Acc=53.7, F-score=54), maximum velocity (VM, Acc=52.6, F-score=52.4) and jerk (JM, Acc=50.7, F-score=50.5), average variability (VWA, Acc=46.8, F-score=46.9) and maximum variability (VWM, Acc=44.5, F-score=44.6), curvature (CWM, Acc=32.7, F-score=33.2), arc length (AL, Acc=31.3, F-score=31.3), AACs (Acc=28.6, F-score=28.7) and suddenness (S, Acc=33.5, F-score=28.7). It is important to notice that the top 7 features also appear (although in different order) among the features which received the highest F-value within the statistical analysis (see Table II).

From the results it can be seen that although the kinematics properties of the gesture (Jerk, Acceleration, Velocity) allow to

TABLE VI
CLASSIFICATION RESULTS FOR A) PRE-SELECTED FEATURES
COMBINATIONS (UPPER PART OF TABLE), AND B) FIVE TO SIX FEATURES
EXTRACTED WITH FACTOR ANALYSIS (LOWER PART OF TABLE).

Feature	Acc.	F-score	Prec.	Rec.
AM + SMS	61.5	61.3	61.3	61.5
AM + VM	62.2	61.8	61.6	62.2
AM + SMA	64.2	63.8	63.7	64.3
AM + VMS	64.4	64.2	64.1	64.4
AM+SMS+VMS	65.7	64.8	65.1	65.7
AM+VM+SMS	66.6	65.9	65.9	66.6
AM+SMM+SMA+SMS+VMS	67.7	67.3	67.4	67.8
AM+VM+SMS+VMS	71.4	70.8	71.1	71.4
JM+SMM+AACS+AES+S	75.9	75.5	75.5	75.9
JM+SMM+AACS+AES+AACM	80.8	80.6	80.6	80.8
JM+SMM+AACS+AES+AACM+S	82.5	82.4	82.4	82.6

distinguish quite well VFs, the other features, and in particular, the features measuring smoothness and motion variability are also powerful VF discriminators. At the same time, one may see the huge difference in terms of Accuracy and F-score between the results reported in the above paragraph and the Table III. It seems that the combination of different gesture features is needed to ensure good recognition rate. To analyze further this hypothesis, we checked the Accuracy and F-score rates for some combinations of the 7 top input features. In the Table VI we provide the results. In the last step, we performed Principal Factor Analysis on 17 features (using eigenvalue-one criterion and Varimax rotation). Six main components resulted from this analysis:

- 1) Component 1: JM (0.908), VWM (0.886), AM (0.868), VWS (0.837), VM (0.644), AL (0.624) which, among others, regroups features related to peak an action,
- 2) Component 2: SMA (0.945), SMM (0.945), SMS (0.866) related to smoothness of the wrist movement,
- 3) Component 3: AACM (0.919), AACS (0.916) - concerns the shoulder angle,
- 4) Component 4: AED (0.859), AEM (0.854) - regards the elbow angle,
- 5) Component 5: AACA (0.953), AL (0.609),
- 6) Component 6: S (-0.737), CWM (0.662), VWA (-0.509), which principally measures whether the movement contains abrupt changes and irregularities.

When considering 6 input features (one per component), the classification rates raise up to 82.4% (against the 84.4% when all the features are used, and 54.4% when only maximum of acceleration is used). This result, further confirms that the VFs consist of combination of kinematics and trajectory-related characteristics.

VII. CONCLUSION AND FUTURE WORKS

In this paper we investigated which features of a movement can endow it with a vitality form and whether it is possible to recognize such forms automatically. To do so, we generated a novel dataset of a variety of actions executed by professional actors, either with rude and gentle vitality forms or without vitality, but with modulations in peak speed. The analysis of the kinematics properties of the collected actions demonstrated that significant differences exist between the features

characterizing gentle and rude forms, even when applied to very different goal directed actions (such as pointing, offering or dropping an object). More interestingly, we showed that actions endowed with vitality forms differ significantly from actions performed with different speeds. In other words, rude actions are different from fast actions, while gentle actions are different from slow actions. A proper differentiation among actions with rude and gentle VFs and actions simply performed at a natural, faster or slower pace can be achieved only considering the the combination of several features, including kinematics (velocity, jerk) properties and trajectory-based features measuring the movement smoothness, variability, and angles. Standard kinematics features (e.g., velocity peak) were not sufficient to distinguish gentle and slow movements on our dataset.

This paper also presents a first attempt to perform automatic vitality forms recognition. Our results show that classifying the different VFs is possible, reaching up to 87% (F-score). Moreover, it can also generalize to gestures, which are not present in the training set, reaching the average score of 73.5% (F-score). Importantly our classifiers are able to distinguish well rude from fast gestures, and, to lower extent, gentle from slow gestures.

Fig. 5. The proposed methodology. Creation of the dataset (1), identification of the features (2), creation of synthesis models for new gestures conveying vitality forms (3-4); evaluation of the human brain activity (fMRI) during the observation of these new gestures (5).

Some limitations of this study should be also mentioned. Although our dataset is quite big, the gestures were acted only by two actors. This does not allows us to study interpersonal differences. The data corresponding to other VFs should be also collected in future as well as more complex actions including other body parts. Applying more advanced techniques of machine/deep learning to (eventually) improve classification results is out of the scope of this paper. As stated in Section II VFs can be expressed with different modalities. Multimodal approaches will also be addressed in future works. Last but not least, the models presented in this paper are based on MoCap data. Although MoCap ensures the best quality data,

using it in out-of-lab human-agent interaction is not feasible. Thus, we plan to downscale our approach to be used with other data sources, e.g., RGB cameras (similar 2-steps approach was applied in the past, e.g., in [45]).

This work is a part of larger project aiming to provide artificial agents with the capacity to perceive and generate vitality forms. Thus, the knowledge about gentle and rude movement features discovered in this work, will be used, in the next stage of the project, to define synthesis models for artificial agent. We are currently developing such models for the iCub humanoid robot and its virtual simulator (avatar). Using such models, rude or gentle VF will be applied to a MoCap data of a “neutral” gesture that is not present in the dataset. Finally, the validity of the model will be assessed with fMRI by measuring the neural activity elicited during their observation. This last step will be realized using the same approach that was used in [20]. We will compare the neural activity during the observation of human gesture and new robotic gestures assessing similarities and possible differences.

Our study represents the first attempt to assess the spatiotemporal features essential in conveying gentle and rude vitality forms. Additionally, our results indicate that vitality forms are not merely characterized by a velocity/acceleration modulation but by a combination of different spatiotemporal properties. They represent a fundamental step toward enabling artificial agent to understand human attitudes during interaction, so as to properly adapt their behavior to the state of the partner.

Once we are able to develop VF models, several applications are possible. For example, artificial agents endowed with a capacity to generate and understand vitality forms may appear in several scenarios such as assistance of elderly people and hospital care (see, e.g., [47], [48]), serious games and social skills training (see, e.g., [9]), autism therapy (see e.g., [4]), entertainment (e.g., video games), security and surveillance (e.g., [8] and customer services. We believe that the interdisciplinary approach presented in this paper will allow advancement in theory and methods providing new outcomes that will have an impact in different fields ranging from neuroscience to social humanoid robotics and virtual agents’ community.

ACKNOWLEDGMENT

Alessandra Sciutti and Giuseppe Di Cesare are supported by a Starting Grant from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program. G.A. No. 804388, wHiSPER.

REFERENCES

- [1] D. N. Stern, *Forms of vitality exploring dynamic experience in psychology, arts, psychotherapy, and development*. Oxford University Press, 2010.
- [2] —, “Vitality contours: The temporal contour of feelings as a basic unit for constructing the infant’s social experience.” in *Early social cognition: Understanding others in the first months of life*, 1999, pp. 67–80.
- [3] G. Di Cesare, C. Di Dio, M. J. Rochat, C. Sinigaglia, N. Bruschweiler-Stern, D. N. Stern, and G. Rizzolatti, “The neural correlates of ‘vitality form’ recognition: an fMRI study: This work is dedicated to Daniel Stern, whose immeasurable contribution to science has inspired our research.” *Social Cognitive and Affective Neuroscience*, vol. 9, no. 7, pp. 951–960, 06 2013.
- [4] M. J. Rochat, V. Veroni, N. Bruschweiler-Stern, C. Pieraccini, F. Bonnet-Brilhault, C. Barthélmy, J. Malvy, C. Sinigaglia, D. N. Stern, and G. Rizzolatti, “Impaired vitality form recognition in autism,” *Neuropsychologia*, vol. 51, no. 10, pp. 1918–1924, 2013.
- [5] L. Casartelli, A. Cesaro, E. Biffi, G. Campione, L. Villa, M. Massimo, and C. Sinigaglia, “Vitality form expression in autism,” *Scientific Reports*, vol. 10, 10 2020.
- [6] R. Niewiadomski and C. Pelachaud, “Affect expression in ECAs: Application to politeness displays,” *Int. J. Hum.-Comput. Stud.*, vol. 68, no. 11, pp. 851–871, Nov. 2010.
- [7] S. Zojaji, C. Peters, and C. Pelachaud, “Influence of virtual agent politeness behaviors on how users join small conversational groups,” in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, ser. IVA ’20. New York, NY, USA: Association for Computing Machinery, 2020.
- [8] O. Inbar and J. Meyer, “Politeness counts: Perceptions of peacekeeping robots,” *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 232–240, 2019.
- [9] N. Vannini, S. Enz, M. Sapouna, D. Wolke, S. Watson, S. Woods, K. Dautenhahn, L. Hall, A. Paiva, E. André, R. Aylett, and W. Schneider, “‘FearNot!’: a computer-based anti-bullying-programme designed to foster peer intervention,” *European Journal of Psychology of Education*, vol. 26, no. 1, pp. 21–44, 2011.
- [10] G. Di Cesare, C. Di Dio, M. Marchi, and G. Rizzolatti, “Expressing our internal states and understanding those of others,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 33, pp. 10331–10335, 2015.
- [11] K. R. Scherer, “What are emotions? and how can they be measured?” *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.
- [12] A. Camurri, G. Volpe, S. Piana, M. Mancini, R. Niewiadomski, N. Ferrari, and C. Canepa, “The dancer in the eye: Towards a multi-layered computational framework of qualities in movement,” in *Proceedings of the 3rd International Symposium on Movement and Computing*, ser. MOCO ’16. New York, NY, USA: Association for Computing Machinery, 2016.
- [13] R. Niewiadomski, M. Mancini, and S. Piana, “Human and virtual agent expressive gesture quality analysis and synthesis,” in *Coverbal Synchrony in Human-Machine Interaction*, M. Rojc and N. Campbell, Eds. CRC Press, 2013, pp. 269–292.
- [14] G. Varni and M. Mancini, *Movement Expressivity Analysis: From Theory to Computation*. Cham: Springer International Publishing, 2020, pp. 213–233.
- [15] G. Rizzolatti, A. D’Alessio, M. Marchi, and G. Cesare, “The neural bases of tactile vitality forms and their modulation by social context,” *Scientific Reports*, vol. 11, 04 2021.
- [16] G. Di Cesare, V. Cuccio, M. Marchi, A. Sciutti, and G. Rizzolatti, “Communicative And Affective Components in Processing Auditory Vitality Forms: An fMRI Study,” *Cerebral Cortex*, 08 2021, bhab255.
- [17] G. Di Cesare, F. Fasano, A. Errante, M. Marchi, and G. Rizzolatti, “Understanding the internal states of others by listening to action verbs,” *Neuropsychologia*, vol. 89, pp. 172 – 179, 2016.
- [18] G. Di Cesare, M. Marchi, A. Errante, F. Fasano, and G. Rizzolatti, “Mirroring the Social Aspects of Speech and Actions: The Role of the Insula,” *Cerebral Cortex*, vol. 28, no. 4, pp. 1348–1357, 03 2017.
- [19] G. Cesare, F. Vannucci, F. Rea, A. Sciutti, and G. Sandini, “How attitudes generated by humanoid robots shape human brain activity,” *Scientific Reports*, vol. 10, 10 2020.
- [20] F. Vannucci, G. Di Cesare, F. Rea, G. Sandini, and A. Sciutti, “A robot with style: Can robotic attitudes influence human actions?” in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, 2018, pp. 1–6.
- [21] R. Laban and F. C. Lawrence, *Effort*. Macdonald & Evans, 1947.
- [22] B. Ran, S. Tal, T. Rachelle, S. Karen, and S. Assaf, “Multitask learning for laban movement analysis,” in *Proceedings of the 2nd International Workshop on Movement and Computing*, ser. MOCO ’15, 2015, pp. 37–44.
- [23] D. Swaminathan, H. Thornburg, J. Mumford, S. Rajko, J. James, T. Ingalls, E. Campana, G. Qian, P. Sampath, and B. Peng, “A dynamic bayesian approach to computational laban shape quality analysis,” *Advances in Human-Computer Interaction*, vol. 362651, p. 17, 2009.
- [24] A. Samadani, S. Burton, R. Gorbet, and D. Kulic, “Laban effort and shape analysis of affective hand and arm movements,” in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, Sept 2013, pp. 343–348.
- [25] G. Caridakis, A. Raouzaoui, E. Bevacqua, M. Mancini, K. Karpouzis, L. Malatesta, and C. Pelachaud, “Virtual agent multimodal mimicry of

- humans,” *Language Resources and Evaluation*, vol. 41, no. 3, pp. 367–388, 2007.
- [26] R. Niewiadomski, M. Mancini, G. Volpe, and A. Camurri, “Automated detection of impulsive movements in HCI,” in *Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter*, ser. CHIItaly 2015. New York, NY, USA: ACM, 2015, pp. 166–169.
- [27] B. Mazzarino and M. Mancini, “The need for impulsivity & smoothness - improving HCI by qualitatively measuring new high-level human motion features,” 01 2009, pp. 62–67.
- [28] R. Niewiadomski, M. Mancini, S. Piana, P. Alborno, G. Volpe, and A. Camurri, “Low-intrusive recognition of expressive movement qualities,” in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ser. ICMI 2017. New York, NY, USA: ACM, 2017, pp. 230–237.
- [29] 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, Jan 2013.
- [30] M. Karg, A. Samadani, R. Gorbet, K. Kühnlenz, J. Hoey, and D. Kuli?, “Body movements for affective expression: A survey of automatic recognition and generation,” *IEEE Trans. on Affective Computing*, vol. 4, no. 4, pp. 341–359, 2013.
- [31] G. Castellano, S. D. Villalba, and A. Camurri, “Recognising human emotions from body movement and gesture dynamics,” in *ACII*, A. C. R. Paiva, R. Prada, and R. W. Picard, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 71–82.
- [32] S. Piana, A. Staglianò, F. Odone, and A. Camurri, “Adaptive body gesture representation for automatic emotion recognition,” *ACM Trans. on Interactive Intelligent Systems*, vol. 6, no. 1, pp. 6:1–6:31, Mar. 2016.
- [33] C. Beyan, S. Karumuri, G. Volpe, A. Camurri, and R. Niewiadomski, “Modeling multiple temporal scales of full-body movements for emotion classification,” *IEEE Transactions on Affective Computing*, pp. 1–1, 2021.
- [34] M. R. Loghmani, S. Rovetta, and G. Venture, “Emotional intelligence in robots: Recognizing human emotions from daily-life gestures,” in *IEEE Conf. Robotics and Automation (ICRA)*. IEEE, 2017, pp. 1677–1684.
- [35] N. Fourati and C. Pelachaud, “Emilya: Emotional body expression in daily actions database,” in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*. Reykjavik, Iceland: European Language Resources Association (ELRA), May 2014, pp. 3486–3493.
- [36] DMCD, “Dance Motion Capture Database: <http://dancedb.eu/>,” 2021.
- [37] G. Di Cesare, G. Valente, C. Di Dio, E. Ruffaldi, M. Bergamasco, R. Goebel, and G. Rizzolatti, “Vitality forms processing in the insula during action observation: A multivoxel pattern analysis,” *Frontiers in Human Neuroscience*, vol. 10, p. 267, 2016.
- [38] G. Metta, P. Fitzpatrick, and L. Natale, “Yarp: Yet another robot platform,” *International Journal of Advanced Robotic Systems*, vol. 3, no. 1, p. 8, 2006.
- [39] G. Volpe, P. Alborno, A. Camurri, P. Coletta, S. Ghisio, M. Mancini, A. Massari, R. Niewiadomski, S. Piana, and R. Sagoleo, “Designing multimodal interactive systems using EyesWeb XMI,” in *Smart Ecosystems cReation by Visual dESign Workshop*, 2016, pp. 49–56.
- [40] C. Beyan, V.-M. Katsageorgiou, and V. Murino, “Moving as a leader: Detecting emergent leadership in small groups using body pose,” in *Proceedings of the 25th ACM International Conference on Multimedia*, ser. MM ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1425–1433.
- [41] C. Larboulette and S. Gibet, “A review of computable expressive descriptors of human motion,” in *Proceedings of the 2nd International Workshop on Movement and Computing*, ser. MOCO ’15. New York, NY, USA: Association for Computing Machinery, 2015, p. 21–28.
- [42] R. Niewiadomski, K. Kolykhalova, S. Piana, P. Alborno, G. Volpe, and A. Camurri, “Analysis of movement quality in full-body physical activities,” *ACM Trans. on Interactive Intelligent Systems*, vol. 9, no. 1, pp. 1:1–1:20, Feb. 2019.
- [43] S. Saha, S. Datta, A. Konar, and R. Janarthanan, “A study on emotion recognition from body gestures using kinect sensor,” in *2014 International Conference on Communication and Signal Processing*, 2014, pp. 056–060.
- [44] M. Karg, K. Kühnlenz, and M. Buss, “Recognition of affect based on gait patterns,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 4, pp. 1050–1061, 2010.
- [45] R. Niewiadomski, M. Mancini, G. Varni, G. Volpe, and A. Camurri, “Automated laughter detection from full-body movements,” *IEEE Trans. on Human-Machine Systems*, vol. 46, no. 1, pp. 113–123, Feb 2016.
- [46] N. Fourati and C. Pelachaud, “Multi-level classification of emotional body expression,” in *IEEE Int. Conf. Automatic Face and Gesture Recognition (FG)*, vol. 1, 2015, pp. 1–8.
- [47] N. Lee, J. Kim, E. Kim, and O. Kwon, “The influence of politeness behavior on user compliance with social robots in a healthcare service setting,” *International Journal of Social Robotics*, vol. 9, 11 2017.
- [48] S. Hammer, B. Lugin, S. Bogomolov, K. Janowski, and E. André, “Investigating politeness strategies and their persuasiveness for a robotic elderly assistant,” in *Persuasive Technology*, A. Meschtscherjakov, B. De Ruyter, V. Fuchsberger, M. Murer, and M. Tscheligi, Eds. Cham: Springer International Publishing, 2016, pp. 315–326.

Radoslaw Niewiadomski received the PhD degree in Computer Science from the University of Perugia, Italy. His research interests include emotion recognition, nonverbal behavior synthesis and multimodal interaction. He has been involved in several EU research projects, e.g., FP6 CALLAS, FP7 ILHAIRE and H2020 DANCE and co-authored over 75 peer-reviewed conference and journal papers.

Amrita Suresh is a R&D Engineer at Inria, France, performing research in the area of Autonomous Research Robots and their applications. She holds a Bachelor's degree in Computer Science and Engineering from Amrita Vishwa Vidyapeetham (Amrita University), India, and a double Master's degree in Robotics under the EMARO+ program, from Ecole Centrale de Nantes (France) and University of Genoa (Italy). Her research interests include Cognitive Robotics and Human-Robot Interaction.

Alessandra Sciutti is Tenure Track Researcher, head of the CONTACT (COgNiTive Architecture for Collaborative Technologies) unit of the Italian Institute of Technology (IIT). With a background on Bioengineering, she received her Ph.D. in Humanoid Technologies from the University of Genova in 2010. After two research periods in USA and Japan, in 2018 she has been awarded the ERC Starting Grant wHiSPER (www.whisperproject.eu), focused on the investigation of joint perception between humans and robots. She published more than 60 papers in international journals and conferences. She is currently Associate Editor for several journals, among which the International Journal of Social Robotics, the IEEE Transactions on Cognitive and Developmental Systems.

Giuseppe Di Cesare obtained the PhD in Neuroscience from the University of Parma, Italy. He is the lead research scientist on vitality forms expression and recognition. Since August 2019, he is working at IIT (CONTACT lab) where he is carrying out experiments on human-robot interactions involving fMRI and kinematics data collection which is part of the ERC starting grant awarded to Dr. Sciutti (G.A: 804388).