https://doi.org/10.1145/3136755.3136757

Low-Intrusive Recognition of Expressive Movement Qualities

Radoslaw Niewiadomski
CasaPaganini - InfoMus, DIBRIS, University of Genoa
Viale Causa 13, 16145 Genoa, Italy
radoslaw.niewiadomski@dibris.unige.it

Maurizio Mancini
CasaPaganini - InfoMus, DIBRIS, University of Genoa
Viale Causa 13, 16145 Genoa, Italy
maurizio.mancini@unige.it

Stefano Piana
CasaPaganini - InfoMus, DIBRIS, University of Genoa
Viale Causa 13, 16145 Genoa, Italy
steto84@infomus.org

Paolo Alborno
CasaPaganini - InfoMus, DIBRIS, University of Genoa
Viale Causa 13, 16145 Genoa, Italy
paoloalborno@gmail.com

Gualtiero Volpe
CasaPaganini - InfoMus, DIBRIS, University of Genoa
Viale Causa 13, 16145 Genoa, Italy
gualtiero.volpe@unige.it

Antonio Camurri
CasaPaganini - InfoMus, DIBRIS, University of Genoa
Viale Causa 13, 16145 Genoa, Italy
antonio.camurri@unige.it

ABSTRACT

In this paper we present a low-intrusive approach to the detection of expressive full-body movement qualities. We focus on two qualities: Lightness and Fragility and we detect them using the data captured by four wearable devices, two Inertial Movement Units (IMU) and two electromyographs (EMG), placed on the forearms. The work we present in the paper stems from a strict collaboration with expressive movement experts (e.g., contemporary dance choreographers) for defining a vocabulary of basic movement qualities. We recorded 13 dancers performing movements expressing the qualities under investigation. The recordings were next segmented and the perceived level of each quality for each segment was ranked by 5 experts using a 5-points Likert scale. We obtained a dataset of 150 segments of movement expressing Fragility and/or Lightness. In the second part of the paper, we define a set of features on IMU and EMG data and we extract them on the recorded corpus. We finally applied a set of supervised machine learning techniques to classify the segments. The best results for the whole dataset were obtained with a Naive Bayes classifier for Lightness (F-score 0.77), and with a Support Vector Machine classifier for Fragility (F-score 0.77). Our approach can be used in ecological contexts e.g., during artistic performances.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Interaction devices*; Empirical studies in HCI;

KEYWORDS

HCI; expressive qualities; dance; EMG; IMU

ICMI'17, November 13–17, 2017, Glasgow, UK

ACM ISBN 978-1-4503-5543-8/17/11…$15.00
https://doi.org/10.1145/3136755.3136757

ACM Reference Format:
Radoslaw Niewiadomski, Maurizio Mancini, Stefano Piana, Paolo Alborno, Gualtiero Volpe, and Antonio Camurri. 2017. Low-Intrusive Recognition of Expressive Movement Qualities. In *Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17)*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3136755.3136757

1 INTRODUCTION

Expressive quality of movement is one of the most important aspects of the Human-Human communication of high-level messages (e.g., emotional states, social signals) [8]. Expressivity of movement represents the way in which, for example, a gesture is performed. It carries high-level information, such as the emotional intention of the movement, or its social meaning.

The current approach to expressive movement analysis in the scientific community usually consists in defining movement features (e.g., speed, acceleration, direction, energy, and so on) and then to extract and exploit them to define models of higher-level information (e.g., emotional states, social bonds). Recent computational models and analysis techniques were developed to automatically compute and analyze different movement expressive qualities, e.g., [10, 15, 20], see [21] for a more complete review.

Automated analysis of nonverbal expressive qualities of full-body human movement opens a broad range of applications: for example, therapy and rehabilitation, systems and interfaces enabling deeper experience of audio-visual cultural content (e.g., in museums) and in general novel expressive multimodal interfaces.

In this paper we focus on two such expressive qualities: Lightness and Fragility and we propose computational model to distinguish between these expressive qualities using low-intrusive and low-cost sensors. The choice of the expressive qualities to focus on is motivated by collaboration with the artists. Our approach is characterized by a continuous collaboration and cross-fertilization between science and art. That is, besides being grounded on scientific evidence, e.g., from psychology and motor sciences, the movement qualities we analyze also stem from discussion with choreographers and dancers, i.e., the most skilled people in conveying expressivity through movement. Moreover, most of the existing works use high precision motion capture devices to detect the expressive quality. Using MoCap during a live performance: 1) is not practical (becouse a long time is needed to dress dancers, the system needs to be calibrated before use), 2) it may limit the dancers' movements (inertial MoCap), 3) is difficult in public spaces (optical MoCap). Kinect is sensitive to 1) changing light conditions, 2) number of participants, 3) dimensions of the stage. In our work, instead, we present a study aiming at developing a low-intrusive approach to expressive qualities detection, using data from various type of wearable sensors such as Inertial Movement Units (IMUs) and electromyogram (EMG). Such an approach enables ecological validity of data collection and the development of systems that can be used outside the scientific laboratory.

Our approach is structured as follows: first we create a dataset which focuses on two movement qualities: Fragility and Lightness. The dataset is next segmented and annotated by experts who rank the perceived level of the expressive quality in a segment. Then, we define a set of features on IMU and EMG data inspired to these qualities and we extract them on the corpus. Finally, we propose models to classify segments displaying the two qualities using our features and supervised machine learning techniques. The work presented in the paper is part of the 3 years EU ICT Project DANCE[1]. The aim of the project is to translate expressive movement qualities into the auditory domain by means of interactive sonification, allowing blind and non-blind people the participation in the dance. This paper focuses on the algorithms for the analysis of two expressive qualities. Interactive sonification models are out of the scope of this paper.

The paper is organized as follows: Section 2 provides a brief review of previous work on movement analysis, whilst Section 3 describes the background works that inspired this research. Section 4 presents the dataset of dance performances we collected. Section 5 describes the features we extracted from the dataset. Section 6 is dedicated to the statistical analysis performed on the extracted features and Section 7 to the classification models of Fragility and Lightness. Finally, Section 8 presents an application of the algorithms in the artistic performance and Section 9 concludes the paper and outlines possible further work.

2 STATE OF THE ART

2.1 Classification of Movement Qualities

Since a few years, several approaches were proposed for detection and classification of expressive movement qualities inspired by art. Most of them use high-precision motion capture devices to compute one or more expressive qualities. Camurri and colleagues [6, 7] recorded professional dancers performing non-propositional movements with different emotional content. They carried out subjective (i.e., human) and objective (i.e., automated) evaluation of the communicated emotional content and found out that subjects ratings can be explained by a small set of low-level features.

Truong and colleagues [27] used a machine learning approach for gesture recognition based on descriptors inspired by the Laban Movement Analysis (LMA) [16]. More than 80 descriptors were considered in the paper, which were inspired by different Laban qualities. For example, 10 feature vectors (5 for each hand) were proposed for the Laban's Flow component. They were: mean, standard deviation, ratio between the maximal and mean values, number of local maxima, and the relative temporal instant of the global maximum of hand's trajectories jerk. Authors did not try to determine explicitly the underlying Laban qualities, but they applied them to detect gestures. Next, they extracted the descriptors and applied machine learning algorithms (e.g., SVM, RF, and so on) on the Cambridge Gestural Performance Database [11] containing Kinect data of several basic iconic and metaphoric gestures. Classification performance is around 97% (F-score).

Ran and colleagues [24] built a dataset of 550 movement segments captured with a Kinect sensor, and developed machine learning algorithms to detect Laban qualities. They used a large set of descriptors: 100 features related to Laban's qualities and other 6000 describing the Kinect skeleton data. Next, Multitask Learning was applied to 18 non-disjunctive Laban qualities (Effort Actions, Shape Qualities, and Shape Change) obtaining an F-score of 0.6.

Hachimura and colleagues [13] developed a system to identify characteristic poses from data of motion captured dancing movements. The characteristic poses correspond to the following four Laban Movement Analysis (LMA) components: Space, Weight, Shape, and Time. First, they computed four high-level features, each of them addressing one component. Next, by observing the change over time of these feature values, body movements corresponding to the different LMA components were extracted. In the last step, the authors compared the results of automatic analysis with experts annotation.

Swaminathan and colleagues [25] proposed a Bayesian fusion approach for identifying the LMA Shape quality from motion capture data. Their method uses a dynamic Bayesian network (DBN) to process movement features across the body and across time. The averaged results are Recall 94.9% and Precision 83.13%.

Alaoui and colleagues [1] recorded professional dancers performing four different full-body movement categories from Emio Greco's vocabulary: Breathing, Jumping, Expanding, and Reducing. Then, they extracted a set of 6 movement qualities to drive a mass-spring system: Verticality, Extension, Leg opening, Shifting of weight, Periodicity, and Quantity of Motion. By showing subjects the animation of the mass-spring system, they demonstrated that the 6 movement qualities driving the mass-spring system can successfully communicate the 4 target movement qualities.

Kitsikidis and colleagues [14] exploited multiple Kinect depth sensors to compute the body joints position of a dancer performing traditional Greek dances. They proposed a fuzzy logic model to detect the quality of the dancer's performance, considering the evaluation of an expert of traditional Greek dances as baseline.

Regarding other types of input data (i.e., other than motion captured data), Ward and colleagues [28] proposed an exploratory study of electromyography (EMG) signals corresponding to the execution of different expressive Laban's Effort qualities, such as Flow being Free or Bound. EMG devices were placed on the dancer forearms and EMG signal amplitudes were computed from the captured signals. Authors suggest that this setup can be useful for automatic classification of expressive qualities.

[1] http://dance.dibris.unige.it

2.2 Multimodal Analysis of the Movement

Multiple modalities provide complementary information that, if considered as a whole, allow one to detect expressive qualities. For example, respiration is of paramount importance to explain expressive features, and is strongly related to physical activity: previous works analyze respiration in walking and running rhythm [5] or rowing [4]. In this context, Lussu and colleagues [18] proposed a multimodal approach to distinguish between movements displaying three different expressive qualities: fluid, fragmented, and impulsive movements. The approach is based on the Event Synchronization algorithm [23], which is applied to compute the amount of synchronization between two low-level features extracted from the sound of the respiration captured by a standard microphone placed near to the mouth, and the whole body kinetic energy estimated from motion capture data. Results showed that fragmented movements display higher average synchronization than fluid ones.

Lopes [17] proposed a multimodal model for gesture segmentation based on IMU and EMG sensors placed on the forearm. The performance of the mono and multimodal algorithms were evaluated through 60 sequences performed by 6 users. The monomodal approach (IMU only) obtained the best results in regards to the total segmentation error, whereas the multimodal approach was particularly successful for hand motion movements only. A similar combination of IMU and EMG sensors placed on the user forearm was used by Freixo [12] to detect 12 different classes of gestures.

Masurelle and colleagues [19] proposed a multimodal (motion capture, audio) approach to recognize isolated dance steps using Gaussian mixture models (GMM) and hidden Markov models (HMM). The system exploited motion features extracted from 3D sub-trajectories of dancers' body-joints (generated from motion capture data), using principal component analysis (PCA). These sub-trajectories were obtained thanks to a footstep impact detection module (obtained from recordings of piezoelectric sensors on the dance floor). Using HMMs, the system recognized gestures among six possible classes with a classification performance of 74% (F-score).

3 BACKGROUND

Artists from the performing arts - and in particular dancers, actors, and musicians - can contribute with a consolidated tradition since centuries on expressivity: on how to convey expression and emotion to an audience by means of non-verbal full-body movement and gesture. They are therefore an important source of inspiration to HCI researchers to build computational models capable to analyze expressive qualities and too build novel multimodal interfaces for non-verbal full-body interaction. Our approach to expressive movement analysis is an example of such intersection of science and art, where HCI, biomechanics, as well as artistic and humanistic theories complement each other.

The EU ICT Project DANCE is a perfect example of this approach: a successful encounter between science and art applied to expressive movement analysis. In the past 2 years, during several meetings, interviews, and movement recording sessions with the famous contemporary dance choreographer Virgilio Sieni [2], we defined an expressive vocabulary of movement basics allowing a person to communicate, for example, emotional states. Our interaction with the choreographer did not involve studying dance movements in a traditional sense. Instead, we asked him to focus on the movement qualities involved in Human-Human communication in real-life.

We rely our work on the framework for the analysis of expressive content conveyed by full-body movement proposed by [9]. The framework is composed of four layers. The first layer is devoted to capturing and preprocessing data from physical sensors (video, motion capture, audio, or wearable sensors). The second one computes low-level motion features at a small time scale (i.e., observable frame-by-frame), such as kinetic energy or smoothness. The third one segments the flow of movements in a series of single units (or gestures) and computes a set of mid-level features such as fluidity or impulsivity, i.e., complex features that are usually extracted on more than one joint, and require significantly longer temporal intervals to be observed (i.e., between 0.5 s and 5 s). Finally, the fourth layer represents even more abstract concepts such as emotional states of the displayer, social attitudes, user's engagement in a full-body interaction.

For the work reported in this paper, we consider two qualities identified by the choreographer V. Sieni, and belonging to the third layer of the framework: Lightness and Fragility. In the following sections, we provide a brief description of each one of them.

3.1 Lightness

A necessary condition for a Light movement is the presence of Fluidity. Further, a movement expressing Lightness should include at least one of the following characteristics: (i) it should exhibit a low amount of downward vertical acceleration following gravity (in particular on forearms and knees), (ii) each possible downward acceleration should be counterbalanced by an opposite "harmonic" upward movement (simultaneous or consequent); (iii) vertical downward acceleration movements should be finalized on the horizontal plane. An example of a dancer performing Light movements can be seen at: https://youtu.be/5Yk35QgyQ1A

3.2 Fragility

Fragility is defined as a sequence of non-rhythmical upper body cracks and leg releases. It emerges, for example, when moving at the boundary between balance and fall, resulting in short movements with continuous interruption and re-planning of motor plans. The resulting movement is non-predictable, interrupted, and uncertain. An example of a dancer performing Fragile movements can be seen at: https://youtu.be/XcEhc0_uuvA

4 DATASET

4.1 Recordings Protocol

For the purpose of this work, we recorded a dataset of short performances of dancers asked to perform full body movements displaying a requested expressive quality. At the beginning of each session, dancers were given the definitions of the expressive qualities. Next, the dancers were asked to perform an improvised choreography containing movements that, in the opinion of the dancer, expressed convincingly the quality. 13 female dancers participated in the data collection process. They had different dance backgrounds

[2]http://www.virgiliosieni.it

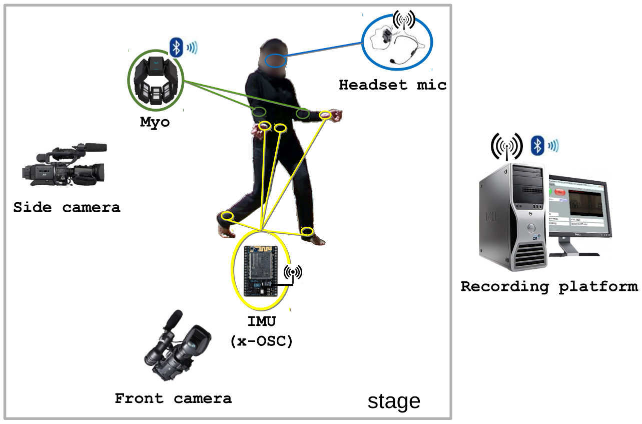

Figure 1: Setup for multimodal recordings

(classic dance, pop, contemporary dance), and different levels of professional experience. They performed five repetitions of each expressive quality, each trial had a duration of 1 minute. All dancers were wearing black clothes.

4.2 Data Streams

We recorded data streams from the following devices:

- 5 IMU sensors (x-OSC[3]) placed on the dancer's body limbs; the data is captured at 50 frames per second; each frame consists of 9 values: (x, y, z) of accelerometer, gyroscope, and magnetometer;
- 2 video cameras (1280x720, at 50fps);
- 2 EMG armband placed on the dancer's forearms (MYO[4]); the data is captured at 50 frames per second. Each armband streams 8 different EMG signals at each frame;
- one wireless microphone (Mono, 48 kHz) placed close to the dancer's nose, recording the sound of breathing;

Figure 1 shows the recording setup. Data was recorded and synchronized using the freely available EyesWeb XMI research platform[5]. Synchronization of data streams is obtained by using SMPTE timecodes, i.e., a standard which is widely used in multimedia content production. In this paper, we use only the data from the 2 IMUs placed on the participant's hands and 2 EMG bracelets placed on the forearms.

4.3 Segmentation

The recorded video streams were evaluated by dance experts and expressive movement analysis experts. For every trial, they identified segments of about 10s each corresponding to a uniform, coherent sequence of movements. End of a segment should correspond to the end of the gesture or phase of the gesture. In a consequence some segments have slightly different duration. For each dancer and each expressive quality, between 5 and 6 segments were chosen. We obtained 150 segments. The details of the segmentation are presented in Table 1.

[3]http://x-io.co.uk/x-osc
[4]https://www.myo.com
[5]http://www.infomus.org/eyesweb_eng.php

Table 1: List of segments

Quality	No. Segments	Mean duration	Total duration
Lightness	77	10.2 s	13 min 6 s
Fragility	73	10.4 s	12 min 41 s
Total	150	10.3 s	25 min 46 s

4.4 Ranking

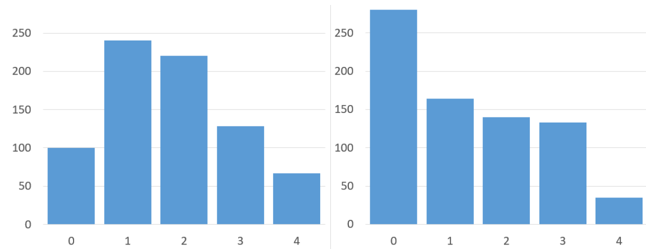

Figure 2: Distribution of the votes in the ranking study: Lightness (left) and Fragility (right)

Five raters watched the 150 segments resulting from the segmentation. They observed each video segment and they were asked to rate the global level of Fragility and Lightness they perceived by using two independent 5-point Likert scales (from 0 to 4).

Exemplary frames of the recordings are displayed in Figure 3. Videos were displayed in random order. The raters did not hear any audio. We blurred the face of the dancer to prevent the rater to identify her and to avoid that facial expressions could affect the ratings. The raters were given the definitions of the expressive qualities and they were explained the features to be computed on the data (see Section 5 for more details on features computation). The latter was made to steer the raters to the measurable aspects of the performances (for example, by letting them know that there is no computation involving feet movements).

We checked the inter-rater reliability between the raters using weighted Cohen κ and Pearson correlation r. The mean pairwise linear weighted Cohen agreement for 5-point scale values are: 0.30 for Lightness and 0.40 for Fragility. The mean correlation values are: 0.46 for Lightness and 0.58 for Fragility.

Figure 3: Two sample frames from the recordings

Next, for each segment we computed the average scores *RankLI* and *RankFR* of the perceived Lightness and Fragility between the

5 raters. As Table 2 shows, many segments received only medium average levels of Lightness and Fragility ranks. Additionally, Lightness scores are better distributed than Fragility scores. Indeed, many segments were perceived as not expressing high Fragility even when the dancers were asked to do it. This might be due to different dance backgrounds of the dancers. In particular, Fragility is a cue that does not appear in classical ballet and thus it might be difficult for some of the dancers to express it in a way that raters can perceive it. Consequently, we decided to define four subsets containing segments of high and low rank for each feature:

- FR_{low} contains 48 segments (out of 150) for which the ranked average Fragility score was below 0.4,
- FR_{high} contains 44 segments (out of 150) for which the ranked average Fragility score was above 2,
- LI_{low} contains 48 segments (out of 150) for which the ranked average Lightness score was below 1.2,
- LI_{high} contains 40 segments (out of 150) for which the ranked average Lightness score was above 2.3.

The choice of the thresholds was made to balance the number of segments in each subset.

5 FEATURES AND DESCRIPTORS

We now describe the multimodal features that we extracted from our dataset. The first two are inspired by the movement qualities described in Section 3 and are computed from two IMU (Inertial Movement Unit) sensors placed on the participant's wrists. The second two are computed from two EMG sensors placed on the participant's forearm. For each feature we extract a set of descriptors (e.g., mean, standard deviation, and so on). The following sections illustrate the features and the corresponding descriptors.

5.1 From IMU

5.1.1 Feature Q_1. Feature Q_1 is inspired by the definition of Lightness reported in Section 3.1. Q_1 is computed in two steps: first, for each IMU sensor i, we compute the ratio W_i between the vertical component of kinetic energy and the total (on the 3-axis) energy. Next we compute:

$$Q_1 = 1 - \frac{\sum_{i=1}^{2} W_i}{2} \tag{1}$$

For the multimodal classification presented in Section 6, the following descriptors are computed: $MEAN_{Q1}$, STD_{Q1}, MIN_{Q1}, MAX_{Q1}.

5.1.2 Feature Q_2. Symmetrically to Q_1, feature Q_2 is inspired by the definition of Fragility reported in Section 3.2. To extract it, we apply the following heuristic: we consider the start and stop

Table 2: Number of segments per rank interval

	Lightness	Fragility
Average rank 1 or less	39	73
Average rank between 1 and 2	60	39
Average rank between 3 and 2	35	32
Average rank between 4 and 3	16	6
Total	150	150

instants of both hands movements; then we measure the synchronization between start/stop instants of both hands (i.e., we check whether both hands start/stop to move simultaneously). If synchronization emerges in one or both hands then we identify this moment as an *Upper body crack*. Finally, we integrate over time the result to obtain Q_2.

The algorithm for Q_2 detection is illustrated in Figure 4. Input consists of the 3-axis linear acceleration of the two inertial sensors attached to the participant's wrists. For each hand, the algorithm detects movement start and stop by looking for acceleration and deceleration peaks. The output of this process are four binary time series, where a 1 or 0 encode whether a peak is detected or nor respectively. Peak synchronization is computed between left/right hand starts and stops events separately by using the Multi-Event Class Synchronization (MECS) algorithm [2] (kernel uniform, and Tau = 10 samples) on a sliding window buffer. If starts or stops events are synchronized, then an *Upper body crack* is detected. Q_2 is finally computed as the integral over time of the output of the upper body detection crack process on a time window. For the multimodal classification presented in Section 6, the following descriptors are computed: $MEAN_{Q2}$, STD_{Q2}, MIN_{Q2}, MAX_{Q2}.

5.2 From EMG

The EMG signal captured on the forearm (8 sensors per armband) was successfully used for gesture recognition in [22, 26]. Recently, it was also used in the study on Laban qualities [28] (see Section 2). Indeed, Fragility segments are composed of movements which require alteration of muscle tension and relaxation phases, so we expect to detect them from such a signal. As previously mentioned in Section 4.2, the MYO armband generates 8 values per frame. Following the literature on EMG signal processing (e.g., [3]) we apply a low-pass filter on the evolution of these 8 values in time. Next, for each data frame d_k, $k = 1..N$ consisting of 8 values x_j^k, $j = 1..8$, we combine the values of all channels using two methods:

$$E_1(d_k) = \frac{\sum_j^8 \|x_j^k\|}{8} \tag{2}$$

$$E_2(d_k) = max_j(x_j^k) \tag{3}$$

For each segment of data we apply 6 descriptors. The first four: mean (MEAN), standard deviation (STD), minimum (MIN), and maximum (MAX) are computed according to their well known equations. Additionally, we also compute Willison Amplitude (WAMP) and Waveform Length (WL) defined as follows:

$$WAMP_{Ei} = \sum_{k=1}^{N} f(\|E_i(d_k) - E_i(d_{k+1})\|), where$$
$$f(x) = \begin{cases} 1 & \text{if } x < threshold \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$WL_{Ei} = \sum_{k=1}^{N} \|E_i(d_{k+1}) - E_i(d_k)\| \tag{5}$$

All the descriptors are computed for each hand separately and next the mean between the descriptors of the two hands is computed.

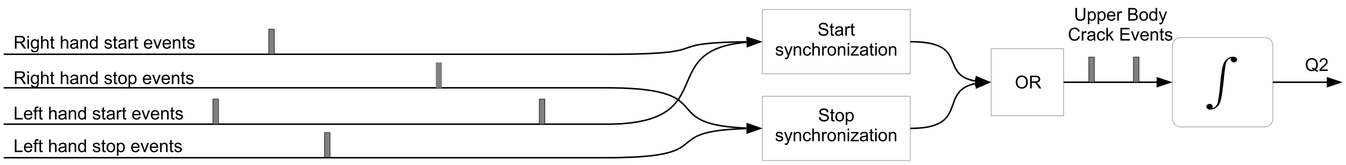

Figure 4: Feature Q_2 extraction algorithm. Acceleration and deceleration phases were extracted on both hands. If synchronization between the two hands peaks is present, then an upper body crack is detected.

6 VALIDATION

The features proposed in the previous section were extracted on the segments in sets FR_{low}, FR_{high}, LI_{low} and LI_{high}. Then, we applied a statistical analysis to check whether there are significant differences between the values of each feature for the two pairs of sets: FR_{low}, FR_{high} and LI_{low}, LI_{high}.

First, the values of all the descriptors were normalized in interval $[0, 1]$. Next, we applied ANOVA with LIGHTNESS (FR_{low} vs. FR_{high}) as independent variable and the descriptors $MEAN_{Q1}$, STD_{Q1}, MAX_{Q1}, MIN_{Q1}, $MEAN_{E1}$, STD_{E1}, $WAMP_{E1}$, WL_{E1}, MAX_{E1}, MIN_{E1}, $MEAN_{E2}$, STD_{E2}, $WAMP_{E2}$, WL_{E2}, MAX_{E2}, MIN_{E2} as dependent variables (see Table 3). As a result, descriptors $MEAN_{Q1}$ ($F(1, 86) = 57.743$, $p < 0.001$), MAX_{Q1} ($F(1, 86) = 4.315$, $p < 0.05$), MIN_{Q1} ($F(1, 86) = 65.102$, $p < 0.001$) had a significantly higher value for segments perceived to express a high Fragility, whilst descriptors STD_{Q1} ($F(1, 86) = 72.865$, $p < 0.001$), $MEAN_{E1}$ ($F(1, 86) = 5.918$, $p < 0.05$), STD_{E1} ($F(1, 86) = 21.383$, $p < 0.001$), WL_{E1} ($F(1, 86) = 4.974$, $p < 0.05$), MAX_{E1} ($F(1, 86) = 29.814$, $p < 0.001$) $MEAN_{E2}$ ($F(1, 86) = 4.748$, $p < 0.05$), STD_{E2} ($F(1, 86) = 12.624$, $p < 0.01$), MAX_{E2} ($F(1, 86) = 11.585$, $p < 0.01$) had significantly lower values.

Similarly, we applied ANOVA with FRAGILITY (FR_{low} vs FR_{high}) as independent variable and the descriptors $MEAN_{Q2}$, STD_{Q2}, MAX_{Q2}, MIN_{Q2}, $MEAN_{E1}$, STD_{E1}, $WAMP_{E1}$, WL_{E1}, MAX_{E1}, MIN_{E1}, $MEAN_{E2}$, STD_{E2}, $WAMP_{E2}$, WL_{E2}, MAX_{E2}, MIN_{E2} as dependent variables (see Table 3).

As a result, descriptors $MEAN_{Q2}$ ($F(1, 90) = 67.368$, $p < 0.001$), STD_{Q2} ($F(1, 90) = 70.368$, $p < 0.001$), MAX_{Q2} ($F(1, 90) = 92.379$, $p < 0.001$), MIN_{Q2} ($F(1, 90) = 20.435$, $p < 0.001$) had significantly higher values for segments perceived to express a high Fragility while descriptors $MEAN_{E1}$ ($F(1, 90) = 3.986$, $p < 0.05$), $WAMP_{E1}$ ($F(1, 90) = 6.778$, $p < 0.05$), MIN_{E1} ($F(1, 90) = 9.733$, $p < 0.01$), $WAMP_{E2}$ ($F(1, 90) = 18.326$, $p < 0.001$), MAX_{E2} ($F(1, 90) = 11.576$, $p < 0.01$) had significantly lower values.

Table 3: Means and standard deviations obtained for each descriptor and subset of the dataset. Significant differences are in bold.

Desc.	*LIlow*	*LIhigh*	*FRlow*	*FRhigh*
$MEAN_{Q1}$	**0.59 (0.24)**	**0.90 (0.09)**	-	-
STD_{Q1}	**0.49 (0.21)**	**0.17 (0.13)**	-	-
MAX_{Q1}	**0.94 (0.19)**	**0.99 (0.0)**	-	-
MIN_{Q1}	**0.39 (0.23)**	**0.77 (0.20)**	-	-
$MEAN_{Q2}$	-	-	**0.05 (0.10)**	**0.43 (0.30)**
STD_{Q2}	-	-	**0.08 (0.14)**	**0.39 (0.21)**
MAX_{Q2}	-	-	**0.09 (0.16)**	**0.50 (0.24)**
MIN_{Q2}	-	-	**0.01 (0.02)**	**0.13 (0.20)**
$MEAN_{E1}$	**0.40 (0.23)**	**0.28 (0.18)**	**0.32 (0.19)**	**0.24 (0.17)**
STD_{E1}	**0.47 (0.24)**	**0.25 (0.19)**	0.31 (0.20)	0.29 (0.18)
MAX_{E1}	**0.51 (0.20)**	**0.28 (0.20)**	0.34 (0.20)	0.36 (0.19)
MIN_{E1}	0.35 (0.20)	0.34 (0.20)	**0.38 (0.20)**	**0.26 (0.15)**
$WAMP_{E1}$	0.74 (0.12)	0.74 (0.16)	**0.76 (0.14)**	**0.68 (0.15)**
WL_{E1}	**0.36 (0.21)**	**0.27 (0.19)**	0.29 (0.18)	0.24 (0.17)
$MEAN_{E2}$	**0.43 (0.26)**	**0.32 (0.22)**	0.35 (0.20)	0.27 (0.23)
STD_{E2}	**0.50 (0.23)**	**0.33 (0.20)**	0.36 (0.18)	0.32 (0.20)
MAX_{E2}	**0.85 (0.20)**	**0.68 (0.26)**	0.74 (0.23)	0.70 (0.28)
MIN_{E2}	0.27 (0.18)	0.27 (0.19)	**0.31 (0.20)**	**0.19 (0.13)**
$WAMP_{E2}$	0.73 (0.19)	0.73 (0.19)	**0.77 (0.15)**	**0.60 (0.23)**
WL_{E2}	0.45 (0.26)	0.35 (0.25)	0.37 (0.23)	0.29 (0.24)

7 CLASSIFICATION

We build two different models: one for Fragility and one for Lightness using different supervised machine learning algorithms and a subset of 11 descriptors out of the 20 described in the previous section: i.e., $MEAN_{Q1}$, STD_{Q1}, MIN_{Q1}, $MEAN_{Q2}$, STD_{Q2}, MAX_{Q2}, MIN_{Q2}, STD_{E1}, MAX_{E1}, MIN_{E2} and $WAMP_{E2}$. We chose the descriptors with correlation score *RankFR* or *RankLI* above 0.3 or below -0.3.

To build the Lightness model we split the dataset into two classes by applying the threshold of 1.6 on *RankLI* (i.e., the median of *RankLI*). As a consequence, we obtain 72 segments expressing Lightness and 78 expressing no or low Lightness. Similarly, we applied the threshold of 1.2 on *RankFR* (i.e., the median value of *RankFR*) to split the segments in those where Fragility was perceived (68 segments) and those where no or low Fragility was observed (82 segments).

For both models, we tested the performance of 3 supervised machine learning algorithms: a Support Vector Machine (SVM) with polynomial kernel, a Random Forest (RF) and a Naive Bayes (NB) classifier.

The averaged performance of each classifier was assessed via a multiple run and Leave-One-Out Method. In our study, we adopted 100 runs. Table 4 shows the performance of each classifier in terms of average Accuracy, Precision, Recall, and F-score.

Additionally, we also ran the same machine learning algorithms on the pairs of sets FR_{low}, FR_{high} and LI_{low}, LI_{high} (that is, on the 88 segments that obtained high or low ranks of *RankFR*) and the 92 segments that obtained high or low ranks of *RankLI*. As expected, performances are higher with Random Forest compared to the results computed on the whole dataset, obtaining an average

Table 4: Average Accuracy, Precision, Recall, F-score for all segments

	Lightness				Fragility			
	Avg. Accuracy	Avg. Recall	Avg. Precision	Avg. F-score	Avg. Accuracy	Avg. Recall	Avg. Precision	Avg. F-score
SVM	0.68	0.65	0.67	0.66	0.79	0.81	0.74	0.77
NB	0.77	0.79	0.74	0.77	0.75	0.59	0.8	0.68
RF	0.75	0.71	0.75	0.73	0.77	0.72	0.75	0.74

F-score of 0.86 for Lightness and of 0.8 for Fragility. Table 5 shows the results.

8 APPLICATION

Algorithms presented in this paper have been exploited in scientific experiments and public events. In particular, they have been recently presented in the context of an artistic project titled *Atlante del Gesto_Genova*[6], by the choreographer V. Sieni. About 150 citizens of Genova (a town located in the north-west of Italy) participated to a series of workshops (from January to March 2017), to gain sensibility about their body expressive movements. The features algorithms presented in this paper have been used during these workshops to help participants to improve their learning curve. The participants' expressive movement features have been translated into sound qualities that they could hear during the rehearsals/performances. The final performances of this scientific/artistic work took place in several historical sites in the town on the 24, 25, and 26 March 2017 (see Figure 5). During the performances, one of the dancers worn IMUs on her hands and legs. The 2 movement features, Fragility and Lightness, were extracted and translated into audio features. Thus, both the audience and the other dancers were able to hear the expressive intention of the sonified dancer.

9 CONCLUSIONS

In this paper, we presented a multimodal dataset of dancers expressing two movement qualities: Lightness and Fragility. Then, we described a set of features to analyze the dataset movement qualities using low-intrusive wearable IMU and EMG sensors. Results of our statistical analysis show that the proposed features permit to distinguish segments according to the degree of Lightness and Fragility perceived by human observers. We also developed two supervised machine learning models to detect Fragility and Lightness. Our models are able to detect the 2 expressive qualities with an F-score of 0.77. To our knowledge, this is the first attempt to combine kinematic and EMG data to detect expressive qualities in human movement. While most of the previous works in this field exploit motion capture systems, our set of features works on a minimal amount of data, captured by only two IMU and EMG sensors placed on the participant's hands and forearms. Another important contribution of this paper is the applied methodology. By collaborating with dancers and choreographers, we were able to precisely define, in terms of kinematic features, two important expressive qualities that have been rarely discussed in the literature so far.

Although the work is inspired by artistic creation and performing arts are thus a natural application area for it, low-intrusive detection of expressive qualities may have several further important applications in everyday activities. Examples include assistive technologies, e.g., rehabilitation and monitoring of patients with motor and cognitive disabilities (e.g., Alzheimer disease), detection of emotions, cognitive states and personality traits from full-body movements (e.g., signs of hesitation and inserurity), edutainment and entertainment systems (e.g., serious games).

In the future, we plan to add other modalities to our model. In more details, we plan to extract some acoustic features from the

Figure 5: The dancers during the performance at Palazzo Reale. One dancer has IMU sensors placed on her hands (see the black straps on her wrists), and her movement qualities are translated into sound qualities in real-time.

[6]https://www.facebook.com/atlantedelgestoGenova

Table 5: Average Accuracy, Precision, Recall, F-score for "best" segments

	Lightness				Fragility			
	Avg. Accuracy	Avg. Recall	Avg. Precision	Avg. F-score	Avg. Accuracy	Avg. Recall	Avg. Precision	Avg. F-score
SVM	0.76	0.73	0.74	0.73	0.76	0.84	0.79	0.81
NB	0.80	0.73	0.81	0.76	0.76	0.84	0.77	0.80
RF	0.86	0.9	0.82	0.86	0.77	0.81	0.78	0.80

dancer's respiration audio signal recorded during the dataset collection process described in this paper. The recent work by Lussu and colleagues [18] shows that it is possible to infer information about expressive quality of movement from such a kind of data. We also plan to evaluate the contribution of each source of the data (IMU, EMG, respiration) to the classification. Last but not least, in the framework of the DANCE Project, we aim to apply the methodology discussed in this paper to extend the vocabulary of qualities we are able to automatically extract and classify.

ACKNOWLEDGMENTS

This research has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement n. 645553 (DANCE). We thank the choreographer Virgilio Sieni for the in-depth discussion and brainstorming on movement qualities, and for the collaboration on the artistic project *Atlante del Gesto_Genova*. We also thank Roberta Canu, Roberta Messa, Federica Loredan, Nicola Ferrari and Corrado Canepa.

REFERENCES

[1] Sarah Fdili Alaoui, Frederic Bevilacqua, and Christian Jacquemin. 2015. Interactive Visuals As Metaphors for Dance Movement Qualities. *ACM Trans. Interact. Intell. Syst.* 5, 3, Article 13 (Sept. 2015), 24 pages.

[2] P. Alborno, G. Volpe, M. Mancini, R. Niewiadomski, S. Piana, and A. Camurri. 2017. The Multi-Event-Class Synchronization (MECS) Algorithm. *Submitted to IEEE Transactions on Human-Machine Systems* (2017).

[3] Zainal Arief, Indra Adji Sulistijono, and Roby Awal Ardiansyah. 2015. Comparison of five time series EMG features extractions using Myo Armband. In *Electronics Symposium (IES), 2015 International.* IEEE, 11–14.

[4] Anthony H Bateman, Alison H McGregor, Anthony MJ Bull, Peter MM Cashman, and Robert C Schroter. 2006. Assessment of the timing of respiration during rowing and its relationship to spinal kinematics. *Biology of Sport* 23, 4 (2006), 353.

[5] Paolo Bernasconi and Jana Kohl. 1993. Analysis of co-ordination between breathing and exercise rhythms in man. *J. Physiol* 471 (1993), 693–706.

[6] Antonio Camurri, Ingrid Lagerlöf, and Gualtiero Volpe. 2003. Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques. *Int. J. Hum.-Comput. Stud.* 59, 1-2 (2003), 213–225.

[7] Antonio Camurri, Barbara Mazzarino, Matteo Ricchetti, Renee Timmers, and Gualtiero Volpe. 2004. Multimodal Analysis of Expressive Gesture in Music and Dance Performances. In *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers*, Antonio Camurri and Gualtiero Volpe (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 20–39.

[8] Antonio Camurri, Barbara Mazzarino, and Gualtiero Volpe. 2004. Expressive interfaces. *Cognition, Technology & Work* 6, 1 (2004), 15–22.

[9] Antonio Camurri, Gualtiero Volpe, Stefano Piana, Maurizio Mancini, Radoslaw Niewiadomski, Nicola Ferrari, and Corrado Canepa. 2016. 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.* ACM, 6.

[10] George Caridakis, Amaryllis Raouzaiou, Elisabetta Bevacqua, Maurizio Mancini, Kostas Karpouzis, Lori Malatesta, and Catherine Pelachaud. 2007. Virtual agent multimodal mimicry of humans. *Language Resources and Evaluation* 41, 3 (2007), 367–388.

[11] Simon Fothergill, Helena M. Mentis, Pushmeet Kohli, and Sebastian Nowozin. 2012. Instructing People for Training Gestural Interactive Systems. ACM, 1737–1746.

[12] Rui Alberto Esteves Freixo. 2015. *Electromyography and inertial sensor-based gesture detection and control.* Master's thesis. Universidade de Porto.

[13] Kozaburo Hachimura, Katsumi Takashina, and Mitsu Yoshimura. 2005. Analysis and evaluation of dancing movement based on LMA. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on.* IEEE, 294–299.

[14] Alexandros Kitsikidis, Kosmas Dimitropoulos, Erdal Yilmaz, Stella Douka, and Nikos Grammalidis. 2014. Multi-sensor Technology and Fuzzy Logic for Dancer's Motion Analysis and Performance Evaluation within a 3D Virtual Environment. In *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access: 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I*, Constantine Stephanidis and Margherita Antona (Eds.). Springer International Publishing, 379–390.

[15] Andrea Kleinsmith and Nadia Bianchi-Berthouze. 2013. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing* 4, 1 (2013), 15–33.

[16] Rudolf Laban and Frederick Charles Lawrence. 1947. *Effort.* Macdonald & Evans.

[17] João Diogo Faria Lopes. 2016. *Gesture Spotting from IMU and EMG Data for Human-Robot Interaction.* Master's thesis. Universidade de Coimbra.

[18] Vincenzo Lussu, Radoslaw Niewiadomski, Gualtiero Volpe, and Antonio Camurri. 2016. Using the Audio Respiration Signal for Multimodal Discrimination of Expressive Movement Qualities. In *Human Behavior Understanding: 7th International Workshop, HBU 2016, Amsterdam, The Netherlands, October 16, 2016, Proceedings*, Mohamed Chetouani, Jeffrey Cohn, and Albert Ali Salah (Eds.). Springer International Publishing, 102–115.

[19] Aymeric Masurelle, Slim Essid, and Gaël Richard. 2013. Multimodal classification of dance movements using body joint trajectories and step sounds. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2013 14th International Workshop on.* IEEE, 1–4.

[20] Michael Neff and Eugene Fiume. 2003. Aesthetic Edits for Character Animation. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '03).* Eurographics Association, 239–244.

[21] Radoslaw Niewiadomski, Maurizio Mancini, and Stefano Piana. 2013. Human and virtual agent expressive gesture quality analysis and synthesis. *Coverbal Synchrony in Human-Machine Interaction* (2013), 269–292.

[22] Prajwal Paudyal, Ayan Banerjee, and Sandeep K.S. Gupta. 2016. SCEPTRE: A Pervasive, Non-Invasive, and Programmable Gesture Recognition Technology. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16).* ACM, New York, NY, USA, 282–293.

[23] Rodrigo Quian Quiroga, Thoma Kreuz, and Peter Grassberger. 2002. Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Phys. Rev. E* 66 (Oct 2002), 041904. Issue 4.

[24] Bernstein Ran, Shafir Tal, Tsachor Rachelle, Studd Karen, and Schuster Assaf. 2015. Multitask Learning for Laban Movement Analysis. In *Proceedings of the 2Nd International Workshop on Movement and Computing (MOCO '15).* 37–44.

[25] Dilip Swaminathan, Harvey Thornburg, Jessica Mumford, Stjepan Rajko, Jodi James, Todd Ingalls, Ellen Campana, Gang Qian, Pavithra Sampath, and Bo Peng. 2009. A Dynamic Bayesian Approach to Computational Laban Shape Quality Analysis. *Advances in Human-Computer Interaction* 362651 (2009), 17.

[26] Mahmoud Tavakoli, Carlo Benussi, and Joao Luis Lourenco. 2017. Single channel surface EMG control of advanced prosthetic hands: A simple, low cost and efficient approach. *Expert Systems with Applications* 79 (2017), 322 – 332.

[27] Arthur Truong, Hugo Boujut, and Titus Zaharia. 2016. Laban descriptors for gesture recognition and emotional analysis. *The Visual Computer* 32, 1 (2016), 83–98.

[28] Nicholas Ward, Miguel Ortiz, Francisco Bernardo, and Atau Tanaka. 2016. Designing and Measuring Gesture Using Laban Movement Analysis and Electromyogram. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16).* ACM, New York, NY, USA, 995–1000.