Journal on Multimodal User Interfaces (2020) 14:1–15
https://doi.org/10.1007/s12193-019-00302-1

ORIGINAL PAPER

The role of respiration audio in multimodal analysis of movement qualities

Vincenzo Lussu[1] · Radoslaw Niewiadomski[1] · Gualtiero Volpe[1] · Antonio Camurri[1]

Received: 8 March 2018 / Accepted: 1 April 2019 / Published online: 11 April 2019

Abstract
In this paper, we explore how the audio respiration signal can contribute to multimodal analysis of movement qualities. Within this aim, we propose two novel techniques which use the audio respiration signal captured by a standard microphone placed near to mouth and supervised machine learning algorithms. The first approach consists of the classification of a set of acoustic features extracted from exhalations of a person performing fluid or fragmented movements. In the second approach, the intrapersonal synchronization between the respiration and kinetic energy of body movements is used to distinguish the same qualities. First, the value of synchronization between modalities is computed using the Event Synchronization algorithm. Next, a set of features, computed from the value of synchronization, is used as an input to machine learning algorithms. Both approaches were applied to the multimodal corpus composed of short performances by three professionals performing fluid and fragmented movements. The total duration of the corpus is about 17 min. The highest *F*-score (0.87) for the first approach was obtained for the binary classification task using Support Vector Machines (SVM-LP). The best result for the same task using the second approach was obtained using Naive Bayes algorithm (*F*-score of 0.72). The results confirm that it is possible to infer information about the movement qualities from respiration audio.

Keywords Movement expressive qualities · Respiration · Intrapersonal synchronization

1 Introduction

Movement expressive qualities describe how a movement is performed [2]. The same movement can be performed with different qualities, e.g., in a fluid, fragmented, hesitant, impulsive, or contracted way. It has been shown that movement qualities might communicate interpersonal relations [29], personality traits [5], cultural background [44], communicative intentions [8] and emotional states [12].

Many researchers [24,40,56] investigated movement qualities and encoded them into categories. Probably the most well known classification of the movement qualities was proposed by Rudolf Laban [27]. The Laban system has four major components: Body, Effort, Shape, and Space. In particular, Effort and Shape are primarily concerned on movement quality. The Effort is defined by 4 bipolar subcomponents: (i) Space denotes relation with the surrounding space; it can be Direct or Indirect; (ii) Weight describes the impact of movement; it can be Strong or Light; (iii) Time corresponds to the urgency of movement; it can be Sudden or Sustained, and (iv) Flow defines the control of movement; it can be Bound or Free. The Shape is characterized by three subcomponents: Shape Flow, Directional, and Shaping/Carving.

Movement qualities are a very relevant aspect of dance, where, e.g., they convey emotion to external observers, and of various sport activities, where they are factors influencing the evaluation of the performance (e.g., in Karate [30]). They also play an important role in rehabilitation (e.g., Parkinson disease and chronic pain [49]), therapy (e.g., autism [39]), and entertainment (e.g., video-games [6]). Several computational models and analysis techniques for assessing and measuring movement qualities have been proposed (see e.g., [32] for a review), as well as algorithms to automatically

This research has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 824160 (EnTimeMent) and 645553 (DANCE).

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12193-019-00302-1) contains supplementary material, which is available to authorized users.

✉ Radoslaw Niewiadomski
radoslaw.niewiadomski@dibris.unige.it

[1] DIBRIS, Universita degli Studi di Genova, Genoa, Italy

detect and compute movement qualities (see Sect. 2.4 for the detailed overview).

In this paper, we explore whether it is possible to contribute to movement qualities recognition by analyzing the audio of respiration. Respiration is of paramount importance for body movement. The respiration pattern might provoke certain visible movements, e.g., in case of laughter [31] or fatigue [25]. The breathing rhythm can be influenced by body movements, e.g., bowing usually corresponds to the expiration phase. Rhythm of respiration is synchronized with rhythmic motoric activities such as running or rowing [21]. Several physical activities such as yoga or tai-chi explicitly connect physical movement to respiration patterns. There exist several techniques to measure the respiration e.g., through respiration belts. In this paper, we collect the respiration data using the standard microphone placed between nose and mouth. It allows us to collect low-intrusively rich information about the human breathing.

To show that it is possible to infer how a person moves from the audio of respiration, we propose two exploratory studies aiming to distinguish fluid and fragmented movements. We intentionally focus on these relatively easy distinguishable and broad movement categories. If our attempt is successful (i.e., audio respiration data provides the sufficient information), in the future, more difficult tasks can be addressed, e.g., classification of more subtle movement qualities e.g., defined by Laban.

In this work dancers were used to collect the multimodal data of full-body movements as they are used to display a huge variety of movement qualities, and they dedicate a lot of effort and time to exercise their expressive vocabulary. Thus, one can expect that various performances by the same dancer, conveying different movement qualities, can provide a solid ground to base our study upon.

The rest of the paper is organized as follows: in Sect. 2, we present existing works on analysis of human movement and of respiration signals; in Sect. 3 we describe the movement qualities we focus on. Section 4 introduces the overview of proposed approaches and Sect. 5 explains the data collection procedure. The two approaches for qualities classification are presented in details in Sects. 6 and 7. We conclude the paper in Sect. 8.

2 State of the art

2.1 Methods for measuring the respiration

In most of the works that consider respiration, data is captured with respiration sensors such as belt-like strips placed on the chest (e.g., [26,58]), or with other dedicated devices. An example of such a device is the CO2100C module by Biopac[1] that measures the quantity of CO_2 in the exhaled air. This sensor is able to detect very slight changes of carbon dioxide concentration levels. Several alternative solutions were proposed (see [17,43] for recent reviews). Folke and colleagues [17] proposed three major categories of measurements for the respiration signal:

- movement, volume, and tissue composition measurements, e.g., transthoracic impedance measured with skin electrodes placed on chest;
- air flow measurements, e.g., nasal thermistors;
- blood gas concentration measurements, e.g., the pulse - oximetry method that measures oxygen saturation in blood.

The choice of measurement device influences what kind of features can be extracted from the respiration data. Boiten at al. [7] distinguished three classes of approaches to process the respiration signals: (1) the volume and timing parameters, (2) the measures regarding the morphology of the breathing curve, (3) the measures reflecting gas exchange. The first group includes features such as: respiration rate (RR), duration of a respiratory cycle or duration of the interval between the phases. Mean inspiratory (or expiratory) flow rate is an example of the second type of features. Finally, features of the third type measure the quantity of gases in exhaled air.

Recently Cho and colleagues [13] proposed to use the low cost thermal camera to track the respiration phases. The approach was based on tracking the nostril of the user and analyzing temperature variations in this face area to infer inhalation and exhalation cycles.

Another approach is to use Inertial Measurement Units (IMUs). In [28] a single IMU sensor placed on the person's abdomen is used to extract the respiration pattern. The raw signal captured with the IMU device was filtered with an adaptive filter based on energy expenditure (EE) to remove frequencies that were not related to respiration depending on the type of physical activity: Low EE (e.g., sitting) Medium EE (e.g., walking), and High EE (e.g., running).

2.2 Measuring respiration from the audio signal

Some researchers analyzed the respiration sounds captured on the chest wall or trachea (see [34] for the review). They focused on detecting different dysfunctions of the respiratory system by comparing the values of acoustic features between healthy people and patients with some respiratory problems and different pathology classification (e.g., [51]). The audio signal was also used to segment the respiration into phases. For instance, Huq and colleagues [22] distinguished

[1] http://www.biopac.com/.

between the respiration phases using the average power and log-variance of the band-pass filtered tracheal breath. In particular, they found the strongest differences between the two respiratory phases in the intervals 300–450 Hz and 800–1000 Hz for average power and log-variance respectively. Similarly, Jin and colleagues [23] segmented breath using tracheal signals through genetic algorithms.

Acoustic features of respiration captured by the microphone placed near mouth and nose area were explored by Song to diagnose the pneumonia in children [50]. Using supervised learning with more than 1000 acoustic features (prosodic, spectral, cepstral features and their first and second-order coefficients) he obtained 92% accuracy for binary classification task: pneumonia vs. non-pneumonia [50].

Pelegrini and Ciceri [36] studied interpersonal breathing coordination during a joint action. In this context, they checked whether breathing sounds convey information about the activity being performed. They proposed multilayer analysis to respiratory behavior during different joint actions composed of temporal (e.g., respiration rate) and acoustic (e.g., spectral centroid) features. The multilayer analysis provided quantitative measurements of respiratory behavior that enabled descriptions and comparisons between conditions and actions showing the differences between different joint actions performed by participants.

Wlodarczak and Heldner [58] studied the communicative functions of respiratory sounds. They found that acoustic intensity of inhalation is the feature that allows one to detect the forthcoming turn-takings. The inhalations that precede long speech are louder than those which occur during no-speech activity or before short backchannel verbal utterances.

In [1], the audio of respiration captured with a microphone placed near the mouth was used to detect the respiration phases. First authors isolated the respiration segments using a Voice Activity Detection (VAD) algorithm based on short time energy (STE). Next, they computed Mel-frequency cepstrum coefficients (MFCC) of respiration segments, and they applied a linear thresholding on MFCC to distinguish between the two respiration phases.

Yahya and colleagues [59] also detected respiration phases in audio data. Again, a VAD algorithm was applied to the audio signal to identify the respiration segments. Then, several low-level audio features extracted from the segments were used with a Support Vector Machine (SVM) classifier to separate the exhilaration segments from the inspiration ones.

Ruinskiy and colleagues [45] aimed to separate respiration segments from voice segments in audio recordings. First, for each participant, they created a respiration template using a mean cepstrogram matrix. Next they measured similarity between the template and an input segment in order to classify the latter as a *breathy* or *not breathy* one.

2.3 Respiration and physical activities

Several works analyzed respiration in sport activities such as walking and running [4,21], and rowing [3]. Respiration data was also used to detect emotions [26]. Bernasconi and Kohl [4] studied the effect of synchronization between respiration and legs movement rhythms for efficiency of physical activities such as running or cycling. They measured synchronization as a percentage of the coincidence between the beginning of a respiration phase and the beginning of a step (or a pedaling cycle). According to their results, the higher synchronization results in higher efficiency and lower consumption of oxygen.

Bateman and colleagues [3] measured synchronization between the start of a respiration phase, and the phase of a stroke in rowing by expert and non-expert rowers. Respiration phases were detected with a nostril thermistor, whereas the stroke phase (1 out of 4) was detected from the spinal kinematics and the force applied to the rowing machine. When the synchronization was higher, the higher stroke rate was observed for expert rowers. Additionally, the most frequently observed pattern was the two breath cycles per stroke.

Schmid and colleagues [47] analyzed synchronization between postural sway and respiration patterns captured with a respiratory belt at chest level. A difference was observed in respiration frequency and amplitude between sitting and standing position.

2.4 Multimodal analysis and detection of movement qualities

Recently Alaoui and colleagues [16] showed that combining positional, dynamic and physiological information allows for a better characterization of different qualities of Laban's Effort than in unimodal recognition systems. In their work, positional data from motion capture system is associated with Space component, the jerk extracted from the accelerometer placed on the wrist is related to Time component, while the muscle activation signal from the EMG sensor is associated with Weight component. Nevertheless, most of the existing works use motion capture data to recognize and measure movement qualities. For instance, Ran and colleagues [42] applied supervised machine learning to detect Laban qualities from Kinect data. For this purpose, they proposed a large set of descriptors composed of 100 features related to Laban's qualities and other 6000 describing the Kinect skeleton data. For example, Suddenness is computed using the acceleration skewness. In the final step, multitask learning was applied to 18 Laban qualities (Effort Actions, Shape Qualities, and Shape Change) resulted in F-score of 0.6.

Hachimura and colleagues [20] developed a system to detect the poses which correspond to four Laban subcomponents: Space, Weight, Shape, and Time and validated their

method as compared to experts annotation. First, they computed four high-level features, each of them addressing one subcomponent. Next, by observing the change over time of these feature values, body movements corresponding to the different Laban's subcomponents were extracted.

Swaminathan and colleagues [52] proposed a Bayesian fusion approach for identifying the Shape component from motion capture data. Their method used a dynamic Bayesian network to process movement. The results are 94.9% for recall and 83.13% for precision.

Truong and colleagues [54] proposed around 80 descriptors inspired by Laban's movement qualities for machine learning based gesture recognition. For example, Weight subcomponent is estimated with 30 descriptors computed by applying 5 operators (mean, standard deviation, maximal amplitude, number of local minima, relative temporal instant of the global minimum value) to the vertical components of the velocity and acceleration of 3 joints (the center of the hip, the left and right hand). The descriptors were extracted from the Kinect data of basic iconic and metaphoric gestures and several supervised classification algorithms were applied obtaining *F*-score around 97%.

Samadani and colleagues [46] 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. For instance, the Weight was estimated by computing the maximum of the sum of the kinetic energy of the moving parts of the body. Similarly, the Shape Directional was computed from the average trajectory curvature. The approach was validated by measuring the correlation between the algorithm values and the expert annotations. The results are up to 81% on Effort components.

Similarly, other researchers proposed several systems to compute different than Laban's movement qualities from the video. With the aim of emotion classification from the full-body movements Glowinski and colleagues [19] extracted movement features such as: Smoothness, Impulsiveness, Kinetic Energy, Spatial Extent. Similar approach was used by Caridakis and colleagues [11], who extracted movement qualities from the video stream in real time with the purpose of facilitating the interaction between humanoid computer interface and human user. For instance, Fluidity was computed as the sum of the variance of the norms of the motion vectors, Power as the first derivative of the motion vector, Spatial Extent as the distance between hands, and the Overall Activity as the sum of the motion vectors.

Regarding using different data sources than motion captured and video data, Silang Maranan and colleagues [48] used one wrist-mounted accelerometer and supervised machine learning to detect eight Basic Effort Actions of Laban's system. In their approach, multiple sliding time windows were used to analyze movement data incrementally by examining it across three different time scales. Therefore, around 400 low-level motion features were extracted from the accelerometer data, which allowed them to train the model with a weighted accuracy between 55 and 91% depending on the type of Action. Ward and colleagues [57] proposed an exploratory study of electromyography (EMG) signals corresponding to the execution of Free and Bound movements. For this purpose, authors computed the amplitudes of EMG signal from the Myo devices which were placed on the dancer forearms. The same two data sources were fused in the work proposed by Niewiadomski and colleagues [33] to compute two movement qualities from the vocabulary of the choreographer Vittorio Sieni.

3 Movement testbed

Our main goal in this paper is to show that the information obtained from the audio respiration signal is useful to compute how a person moves in terms of her expressiveness. We specifically focus on two very different types of movements: namely fluid and fragmented movements. These two movement categories substantially differ in terms of motor planning. *Fluid movements* are continuous, smooth and harmonious performances of a global (i.e., involving the whole body) motor plan, and without interruptions [37]. *Fragmented movements* are characterized by several abrupt interruptions and re-planning motor strategies.

Fluid and fragmented movements are present in dance context. Fluidity is the fundamental quality e.g., for classical ballet, while Fragmented movements are a part of the expressive vocabularies of the many contemporary choreographers, e.g., Sagi Gross Company[2].

Examples of fluid and fragmented movements can be seen in the video attached to this article as Supplementary Material.

Although neither fluid nor fragmented movements appear in Laban's terminology, there are several reasons to focus on them in our explanatory study. First, in work, Vaessen and colleagues [55] search for distinct brain responses in fMRI data to the visual stimuli of full body movements, which differ in terms of motor planning. Second, several researchers observed spontaneous synchronization between the full-body movements and respiratory rhythms [4,14,35]. This phenomena is often explained with the concept of entrainment, i.e., by a temporal locking process in which one system's motion or signal frequency entrains the frequency of another system [53]. Motivating from these studies, we expect that differences in motor planning and its execution

[2] www.grossdancecompany.com.

might also influence the respiration patterns and intrapersonal synchronization of respiration and body movements. Third, we would like to recall that, because of the differences in terms of motor planning mentioned above, it is impossible for any movement to be fluid and fragmented at the same time (although a movement can be neither fluid nor fragmented). This important property allow us to perform binary discrimination.

4 Overview of the approach

Common approach to recognize movement qualities is using high-precision motion capture systems, then extracting features, and applying classification algorithms that automatically discriminate different qualities (see Sect. 2.4). The motion capture systems are, however, intrusive as they require a set of sensors or markers to be worn. Often they also require calibration which is difficult in a dynamically changing environment such as during the artistic performance. Additionally, the high cost of the technology and long post-processing are other important shortcomings.

In this paper, we explore the audio signal as a source of breathing data. It can be recorded with cheaper, low-intrusive, yet portable and easy-to-use devices. This approach is appropriate to capture e.g., dancers' or athletes' respiration patterns, because they usually do not speak during a performance, but they move a lot and cannot wear cumbersome devices.

Herein, we propose two methods (see Fig. 1):

- unimodal approach; in the case, only audio data is processed to extract the acoustic features; next they are used to train supervised machine learning algorithms for the binary classification,
- multimodal approach; two low-level features are extracted such that one of them is from motion capture data, and the other is from audio data. The degree of synchronization between these two features is measured using Event Synchronization algorithm [41]; the degree of multimodal synchronization permits to discriminate between the two qualities.

It is worth to highlight the differences between two methods. The first one uses only the audio of respiration and computing some acoustic features but this computation may require more computational power. The second approach is based on very simple features that can be easily computed in real-time even on mobile devices, but it requires exact synchronization of data coming from different sensors. Additionally, the first method uses relatively rich information, and thus, we expect that it provides high effectiveness for the discrimination task. On the contrary, the second approach uses only a small piece of respiration information and by using this we want to see whether it is still possible to infer any knowledge about the quality of corresponding movement. Obviously, in the second case, we do not expect comparable Effectiveness from our algorithm.

5 Experimental setup and the data collection

For the purpose of this work, we collected a set of short trials of dancers performing whole body movements with a requested movement quality. Each trial had a duration of 1.5–2 min. At the beginning of each session, dancers were given definitions of the movement quality by means of textual images. More details on the recording procedure is available in [38]. The dancers were asked to perform: (i) an improvised movements that, in their opinion, express the quality convincingly, as well as (ii) several repetitions of predefined sequences of movements by focusing on the given movement quality.

We recorded multimodal data using (i) a Qualisys motion capture system, tracking markers at 100 frames per second; resulting data consists of the 3D positions of 60 markers; (ii) one wireless microphone (mono, 48 kHz) placed close to the dancer's mouth, recording the sound of respiration; (iii) 2 video cameras (1280 × 720, at 50 fps).

The audio signal was recorded by a microphone with a windproof mechanical filter positioned about 2 cm from the

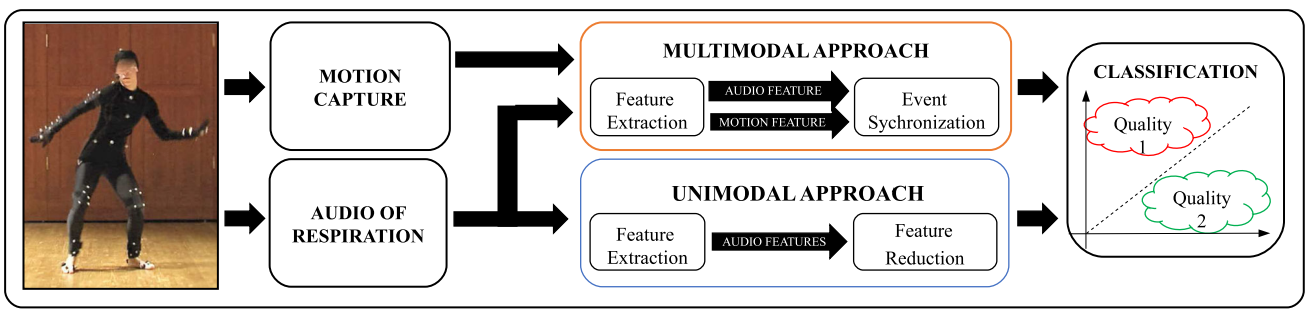

Fig. 1 Two approaches for the classification of the movement qualities from the audio signal of respiration

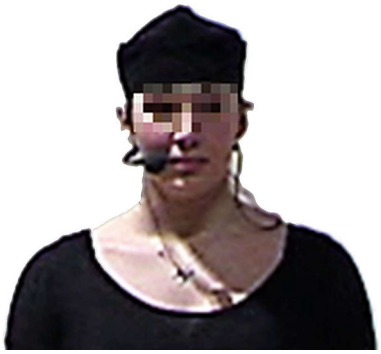

Fig. 2 Correct position of the microphone

Table 1 The quantity and duration of fragmented and fluid episodes

Class	Number	Total duration (min)	Mean (s)	SD (s)
Fragmented	28	7.3	16.3	9.5
Fluid	39	10.1	15.2	15
All	67	17.4	15.6	12.9

nose and mouth (see Fig. 2), ensuring stability of the bow on the dancer's head.

The freely available EyesWeb XMI platform, developed at InfoMus Lab, University of Genoa[3] was used to synchronize recordings and to analyze of the multimodal data.

Motion capture data was cleaned, missing data was filled using linear and polynomial interpolation. 48 kHz audio signals have been pre-processed by applying a high pass filter with a frequency of 200 Hz, as the breath has a bandwidth of between 200 and 2000 Hz [18].

Each quality was performed by three dancers. Next, an expert (by watching just a video) selected from the whole recordings episodes in which dancers have better interpreted one or the other quality. Thus, segmentation was based not only on the dancer's expressive intention, but also on the observer's perception regarding the movement quality. 67 episodes, which are 17.4 min in total, were selected (see Table 1 for details).

6 Classification of fluid and fragmented movements from the unimodal data

The unimodal approach consists of the following steps. First, we automatically extract exhalation phases from the audio corpus. As a result, each exhalation phase becomes an individual segment. We focus on the exhalation signal only since, as it can be seen in Fig. 3, it has a higher spectral energy and generally has a better noise signal ratio (SNR) [1]. Next, we extract Mel-Frequency Cepstral Coefficients (MFCCs), i.e.,

[3] www.infomus.org.

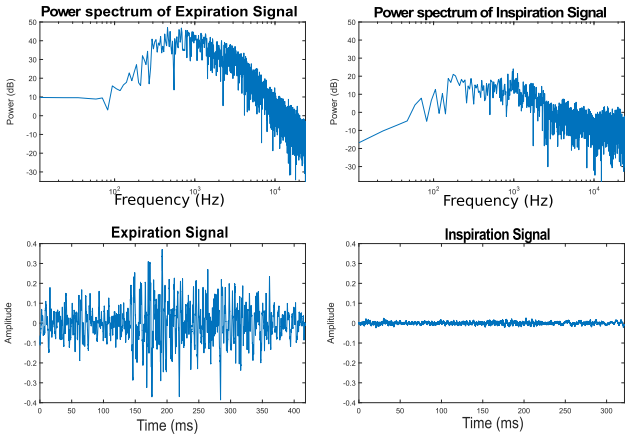

Fig. 3 Top: inhalation and exhalation signals during movement; bottom: the corresponding audio signals

the coefficients that define Mel-Frequency Cepstrum (MFC). The MFC is the representation of the short-term power spectrum of a sound, based on the Mel scale, which approximates the human auditory system.

Consequently, we create four datasets by applying techniques of feature reduction. In order to find the best classification method we train eight classifiers on 2 out of 4 datasets. In the final step the best classification algorithm (in terms of F-score) is applied to all four datasets.

6.1 Data processing

The input to our model is a single exhalation. To obtain exhalations we performed automatic segmentation of the data by modifying the algorithm proposed by Ruinskiy and colleagues [45]. The original algorithm was created to separate respiration segments from voice segments in audio recordings. We adapted this approach to extract the exhalation phases. For this purpose we built an exhalation template (i.e., mean cepstrogram matrix) from the manually annotated one trial of respiration by the Dancer 1 (the total duration of the annotated material was 6.2 s). Next, we applied the template to 67 episodes. As a result, we obtained 467 exhalation segments of total duration of 390.4 s, of which 232.03 s corresponds to fluid movements and 158.37 s corresponds to fragmented movements (see Table 2).

A Mann-Whitney test indicated that there is no significant difference in the durations of fluid and fragmented segments ($U = 25590.5$, $p = .572$). Thus the fluid and fragmented segments cannot be distinguished by considering the durations of their exhalations (Table 3).

The number of fluid and fragmented segments is similar for all three dancers with the small prevalence of fluid over fragmented segments (54, 64 and 60% of fluid movements).

Table 2 The duration of fluid and fragmented segments (only exhalation phase)

Class	Total (s)	Mean (s)	SD (s)
Fragmented	158.37	0.824	0.424
Fluid	232.03	0.844	0.558
All	390.40	0.836	0.507

Table 3 Average and standard deviations of *F*-score and Accuracy obtained by SVM-LP

Dataset	*F*-score	Accuracy
$A_{\text{F-test}}$	0.8689 (0.021)	0.8403 (0.027)
A_{PCA}	0.8571 (0.019)	0.8273 (0.023)
$C_{\text{F-test}}$	0.8558 (0.027)	0.8272 (0.032)
C_{PCA}	0.8201 (0.017)	0.7840 (0.023)

6.2 Feature extraction and reduction

Let us introduce three indexes i, j and k:

- i: index of a segment, i = 1,…, 467,
- j: index of a frame, j = 1,…, N_i,
- k: index of a MFCC coefficient, k = 1,…, 26,

where N_i is the number of audio frames in the segment i (and it varies between segments).

For each exhalation segment, an MFCC matrix is created. We define the MFCC matrix of the i-th segment as $M_i = [m^i_{j,k}]$ with $j = 1, \ldots N_i$ and $k = 1, \ldots 26$. Each element of the matrix corresponds to one audio frame of the exhalation segment of a duration 10 ms. So, duration of the i-th exhalation is $10 \times N_i$ ms and m^i_{jk} is k-th coefficient MFCC of j-th frame of i-th segment.

Next, we reduce the dimensionality of the matrix M_i. For thus purpose we use ten aggregation operators $\Phi_0 - \Phi_9$: Φ_0-Mean, Φ_1-Standard Deviation, Φ_2-Skewness, Φ_3-Minimum, Φ_4-Maximum, Φ_5-Range, Φ_6-Kurtosis, Φ_7-Zero Crossing Rate (ZCR), Φ_8-Linear Trend and Φ_9-Median.

We create two different feature sets:

- feature set A is described by a matrix $\mathbb{A}$ of dimensions 467×130 where each row corresponds to 1 segment and each column contains a result of the application of one aggregation operator between $\Phi_0 - \Phi_9$ on N_i values (i.e., all audio frames) of the coefficient k (where $k = 1, \ldots 13$). Each operator is applied on thirteen MFCC coefficients of the single segment, so we have 10×13 values for each segment. More precisely, an element of the matrix $\mathbb{A}$ is computed according to the formula:

$$a_{i,h} = \underset{j=1,\ldots N_i}{\boldsymbol{\Phi}^{\left\lfloor \frac{h}{13} \right\rfloor}}(m^i_{j,k}) \tag{1}$$

  with $h = 0, \ldots, 129$ and $k = h \mod 13$.
- feature set C is described by a matrix $\mathbb{C}$ of dimensions 467×30 where each row corresponds to 1 segment and each column is an aggregation of the MFCC coefficients. More precisely each element in the matrix $\mathbb{C}$ is computed as follows:

$$c_{i,h} = \begin{cases} \underset{j=1,\ldots N_i}{\Phi^1}(\underset{k=1,\ldots 26}{\Phi^{h-1}}(m^i_{j,k})) & if \quad h < 10 \\ \underset{j=1,\ldots N_i}{\Phi^3}(\underset{k=1,\ldots 26}{\Phi^{h-11}}(m^i_{j,k})) & if \quad 10 \leq h < 20 \\ \underset{j=1,\ldots N_i}{\Phi^4}(\underset{k=1,\ldots 26}{\Phi^{h-21}}(m^i_{j,k})) & if \quad 20 \leq h < 30 \end{cases} \tag{2}$$

  with $k = 1, \ldots 26$ and $h = 0, \ldots 29$.

To avoid the problem of overfitting two standard approaches for feature reduction were applied on $\mathbb{A}$ and $\mathbb{C}$:

- F-test (20 best features for all the dancers),
- Principal Component Analysis (PCA; with 95% of total variance explained).

By applying two features reduction techniques on two matrices $\mathbb{A}$ and $\mathbb{C}$ we obtain four different datasets. Let us introduce the notation X_y where X is the feature set (A or C) and Y is reduction method applied on the feature set (F-test or PCA).

Finally, we performed the exploratory analysis of the four datasets using unsupervised clustering. K-means was applied, and confusion matrices[4] and Cohen κ values were computed for each dataset. The highest Cohen κ was obtained for C_{PCA} ($\kappa = 0.48$), and the second best result was observed for A_{F-test} ($\kappa = 0.4$). Next, we compared the results of two feature reduction approaches. This showed that A_{F-test} had numerically better result than A_{C-test} ($\kappa = 0.28$) while C_{PCA} had numerically better result than A_{PCA} ($\kappa = -0.04$). Therefore, in next Section, we focus on only two best datasets: A_{F-test} and C_{PCA}.

6.3 Classification

In the first step, eight algorithms were tested: CART, Random Forest (RF), ADA, LDA, Naive Bayes (NB), Neural Network (NN), SVM with Gaussian Radial Basis Function (SVM-G)

[4] While calculating confusion matrices, the predicted class label of each cluster was taken according to the majority of the samples real labels.

and SVM with Laplacian RBF (SVM-LP) on A_{F-test} and C_{PCA}.

6.3.1 Comparison of eight classifiers on two datasets

Figure 4 shows the training process. Each dataset was randomly divided into 2 parts: the training set (70%) and the testing set (30%). During the training phase the K-fold algorithm was used with $K = 8$ (inner loop). Next, each classifier was evaluated using the testing set. The same procedure was repeated 10 times: each time the training and testing sets were chosen randomly (outer loop). Accuracy, F-score, Precision, Recall were computed for each iteration of outer loop. In the last step mean and standard deviation of the Accuracy, F-score, Precision, Recall were computed on 10 iterations of the procedure. The corresponding results are presented in Fig. 5, Tables 4 and 5.

To check whether there are significant differences between 8 machine learning algorithms within the dataset we used ANOVA test. Only significant results are listed below. For the post-hoc tests we used Bonferroni correction.

Given A_{F-TEST} ANOVA showed significant difference between the classifiers, $F(7, 72) = 9.4815$, $p < .001$. Using F-score results post-hoc tests showed that:

- CART performed significantly worse than SVM-G, RF, NB, SVM-LP ($p < .05$),
- ADA performed significantly worse than SVM-LP ($p < .05$),
- LDA performed significantly worse than SVM-LP, NN, RF, SVM-G ($p < .05$).

Given C_{PCA} ANOVA did not show significant differences between the classifiers $F(7, 72) = 1.073$, $p = 0.3896$.

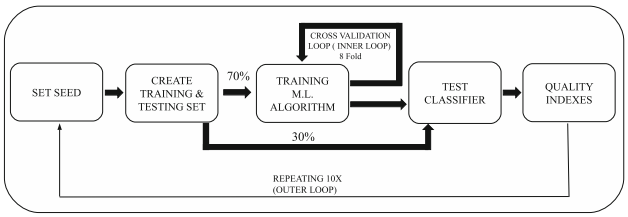

Fig. 4 The overview of machine learning procedure

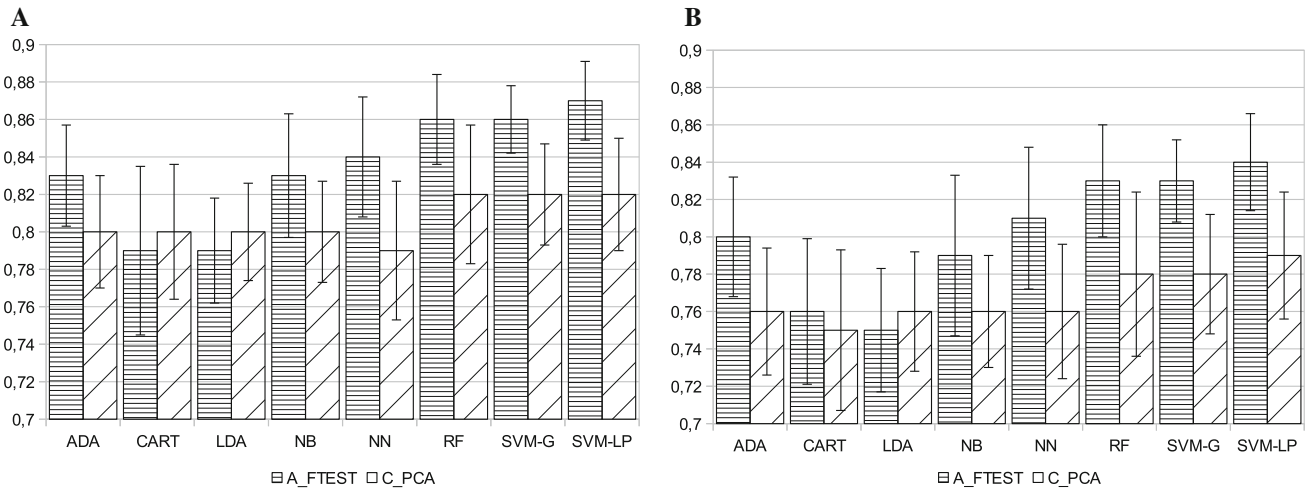

Fig. 5 Average values of the quality indices calculated for each algorithm on the A_{F-TEST} and C_{PCA} datasets: **a** F-score, **b** Accuracy

Table 4 Average values of the quality indices calculated for each algorithm on the A_{F-test} dataset

Algorithm	Accuracy	F-score	Precision	Recall
ADA	0.80 (0.032)	0.83 (0.027)	0.85 (0.046)	0.82 (0.033)
CART	0.76 (0.039)	0.79 (0.045)	0.80 (0.099)	0.80 (0.037)
LDA	0.75 (0.033)	0.79 (0.028)	0.81 (0.044)	0.79 (0.047)
NB	0.79 (0.043)	0.83 (0.033)	0.86 (0.038)	0.80 (0.042)
NN	0.81 (0.038)	0.84 (0.032)	0.83 (0.043)	0.85 (0.045)
RF	0.83 (0.030)	0.86 (0.024)	0.88 (0.035)	0.84 (0.032)
SVM-G	0.83 (0.022)	0.86 (0.018)	0.89 (0.026)	0.84 (0.025)
SVM-LP	0.84 (0.026)	0.87 (0.021)	0.89 (0.029)	0.85 (0.028)

Table 5 Average values of the quality indices calculated for each algorithm on the C_{PCA} dataset

Algorithm	Accuracy	*F*-score	Precision	Recall
ADA	0.76 (0.034)	0.80 (0.030)	0.83 (0.039)	0.78 (0.026)
CART	0.75 (0.043)	0.80 (0.036)	0.85 (0.060)	0.76 (0.038)
LDA	0.76 (0.032)	0.80 (0.026)	0.84 (0.037)	0.76 (0.027)
NB	0.76 (0.030)	0.80 (0.027)	0.84 (0.050)	0.76 (0.023)
NN	0.76 (0.036)	0.79 (0.037)	0.80 (0.050)	0.79 (0.040)
RF	0.78 (0.044)	0.82 (0.037)	0.86 (0.047)	0.78 (0.030)
SVM-G	0.78 (0.032)	0.82 (0.027)	0.85 (0.034)	0.79 (0.026)
SVM-LP	0.79 (0.034)	0.82 (0.030)	0.83 (0.041)	0.81 (0.026)

6.3.2 Comparison of four datasets

Next we created classifiers using the SVM algorithm with Laplacian Kernel for four datasets: A_{F-TEST}, C_{F-TEST}, A_{PCA} and C_{PCA}. This algorithm was chosen as it performed the best in the previous section (see Table 4). The corresponding results are presented in Table 3.

We checked whether there are significant differences between four datasets by applying ANOVA on results of each iteration of training procedure. A significant main effect of dataset was observed for the *F*-score, $F(3, 36) = 9.928$, $p < .001$. Post hoc comparisons with Bonferoni correction showed that *F*-score of C_{PCA} was significantly lower than *F*-score of A_{F-test} ($p < .001$), A_{PCA} ($p < .005$) and C_{F-test} ($p < .005$).

6.4 Conclusion

In this section, we showed that it is possible to distinguish fluid and fragmented movements from the audio of respiration with the Accuracy up to 84% and *F*-score up to 87%. We also compared two different feature reduction techniques and two different features sets for the binary classification task. The best result in terms of *F*-score (87%) was obtained with SVM and 20 features computed from thirteen MFCC coefficients. Similar result was observed when 11 features, which were obtained after applying PCA on the same set of MFCC-based features, were used (*F*-score 86%).

Regarding the differences between classifiers the only significant performance drop was observed in the case of decision tree based algorithms such as ADA and CART. Regarding the performance difference of datasets, unsurprisingly the datasets extracted from matrix $\mathbb{A}$ performed slightly better. It is worth to recall that the initial matrix $\mathbb{A}$ is 4 times more bigger than the initial matrix $\mathbb{C}$, the solution based on matrix $\mathbb{C}$ was only 1% worse than the best solution (see Table 3).

Some shortcomings of this study: first, the data of only three dancers were used. Consequently, we could not validate the classifiers with one-subject-out method. Second, the inhalation data was excluded from the analysis. Possible extensions include checking whether inhalation also brings some useful information for the discrimination task. Third, in the future, other audio features such as spectral centroid will be extracted.

7 Classification of fluid and fragmented movements from the multimodal data

Our multimodal approach is based on hypothesis (H1) that different degree of synchronization can be observed for movements performed with different movement qualities. Our intuition is that if the fluid and fragmented movements differ in terms of motor planning (see Sect. 3) also corresponding respiration patterns may differ.

Our approach is as follows (Fig. 6): from the synchronized recordings we extract one audio feature: the energy of the audio signal, and one movement feature: the kinetic energy of the whole body movement. These features were chosen as they can be easily computed in real-time. In the second step, we define events to be extracted from the time-series of the features values and then we apply the Event Synchronization algorithm [41] to compute the amount of synchronization between them. Next, to check the hypothesis H1 we compare the synchronization degree for two qualities. If the hypothesis H1 is confirmed we build classifiers to distinguish between two qualities.

7.1 Data processing

For the purpose of this study the corpus described in Sect. 5 was used. 9 out of 67 episodes were excluded because of the signal synchronization problems. The remaining episodes have duration between 4 and 85 s. To obtain the segments of comparable duration we split episodes by applying the following procedure:

– episodes were split if they were at least twice longer than the smallest one,

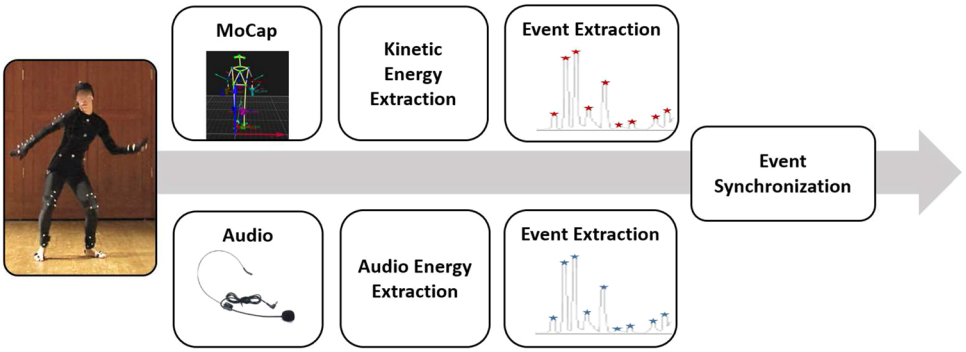

Fig. 6 Block diagram of the analysis procedure. Event Synchronization takes as input events detected in the time-series of (1) energy of the respiration audio signal and (2) kinetic energy from motion capture data

– episodes were split into the segments of the same duration (whether possible).

Consequently, we obtained 192 segments belonging into two classes:

– Fluid Movements Set (*FluidMS*) consisting of 102 segments (average segment duration 4.77 s, sd = 0.70 s);
– Fragmented Movements Set (*FragMS*) consisting of 90 segments (average segment duration 4.193 s, sd = 0.66 s).

7.2 Feature extraction

The audio signal was split in frames of 1920 samples. To synchronize the motion capture data with the audio signal, the former was undersampled at 25 fps. Next, body and audio features were computed separately at this sampling rate.

7.2.1 Motion data

Motion data was used to compute kinetic energy. This feature was computed in two stages: first, 17 markers from the initial set of 60 were used to compute the instantaneous kinetic energy frame-by-frame. The velocities of single body markers contribute to the instantaneous kinetic energy according to the relative weight of the corresponding body parts as retrieved in anthropometric tables [15]. In the second step, the envelope of the instantaneous kinetic energy was extracted using an 8-frames buffer.

7.2.2 Respiration audio

The instantaneous energy of the audio signal was computed using Root Mean Square (RMS). This returns one value for every input frame. Next, we extracted the envelope of the instantaneous audio energy using an 8-frames buffer.

7.3 Synchronization computation

To compute the degree of synchronization we use the Event Synchronization (ES) algorithm [41]. It is used to measure synchronization between two time series in which some events are identified. Let us consider two time-series of features: x_1 and x_2. For each time-series x_i let us define t^{x_i} as the time occurrences of events in x_i. Thus, $t_j^{x_i}$ is the time of the j-th event in time-series x_i. Let m_{x_i} be the number of events in x_i. Then, the amount of synchronization Q^τ is computed as:

$$Q^\tau = \frac{c^\tau(x_1|x_2) + c^\tau(x_2|x_1)}{\sqrt{m_{x_1} m_{x_2}}} \tag{3}$$

where

$$c^\tau(x_1|x_2) = \sum_{i=1}^{m_{x_1}} \sum_{j=1}^{m_{x_2}} J_{ij}^\tau \tag{4}$$

and

$$J_{ij}^\tau = \begin{cases} 1 & if \quad 0 < t_i{}^{x_1} - t_j{}^{x_2} < \tau \\ 1/2 & if \quad t_i{}^{x_1} = t_j{}^{x_2} \\ 0 & otherwise \end{cases} \tag{5}$$

τ defines the length of the synchronization window. Thus, events contribute to the overall amount of synchronization, only if they occur in a τ-long window.

In order to apply the ES algorithm to our data, two steps were needed: (i) defining and retrieving events in two time-series, and (ii) tuning the parameters of the ES algorithm.

7.3.1 Events definition

We defined events as the peaks (local maxima) of kinetic and audio energy. To extract peaks, we applied a peak detector algorithm that computes the position of peaks in an N-size buffer, given a threshold α defining the minimal relative "altitude" of a peak. That is, at time p, the local maximum x_p

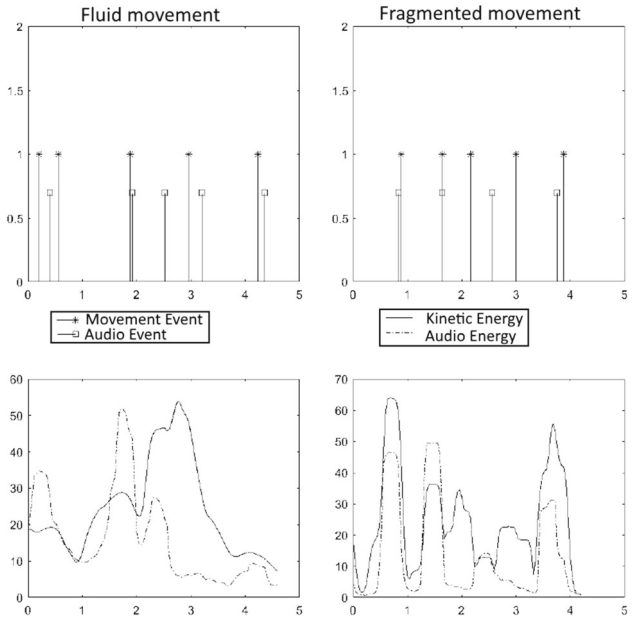

Fig. 7 Excerpts of the two time-series of energy (audio energy and kinetic energy), representing an example of fluid and fragmented movement respectively (lower panel), and the events extracted from the two time-series and provided as input to the ES algorithm (upper panel)

is considered a peak if the preceding and the following local maxima x_i and x_j are such that $x_i + \alpha < x_p$ and $x_j + \alpha < x_p$, $i < p < j$, and there is no other local maximum x_k, such that $i < k < j$. We empirically chose the buffer size to be 10 frames (corresponding to 400 ms) and $\alpha = 0.4465$. Figure 7 shows excerpts of the two time-series, representing an example of fluid and fragmented movement respectively, and the events the peak detector extracted.

7.3.2 Algorithm tuning

At each execution, the ES algorithm works on a sliding window of the data and it computes one value – the amount of synchronization Q^τ. In our case, the value of ES is reset at every sliding window. Thus, the past values of ES do not affect the current output. The algorithm has two parameters: the size of the sliding window dim_{sw} and τ. The size of the sliding window was set to 20 samples (corresponding to 800 ms at 25 fps). This value was chosen as the breath frequency of a moving human is between 35 and 45 cycles per minute. Thus, 800 ms corresponds to half of one breath. We analyzed multimodal synchronization with all τ in interval $[4, dim_{sw} * 0.5]$ (i.e., not higher than half of the size of the sliding window dim_{sw}).

7.4 Data analysis

We utilized datasets *FluidMS* and *FragMS* to test hypothesis (H1). For each segment and each considered value of τ, we computed the average value ($Avg\,Q^\tau$) of the amount of synchronization Q^τ on the whole segment. Next, we computed the mean and standard deviation of $Avg\,Q^\tau$ separately for all fluid and fragmented segments (see Table 6).

To test the differences between the amount of synchronization in the segments of *FluidMS* and *FragMS* we applied Mann-Whitney test on values of $Avg\,Q^\tau$. Similar results were obtained for all the tested τ. A significant effect of $Quality$ for $\tau = 4$ (two tailed, $U = 3598$, $p < .01$), $\tau = 6$ (two tailed, $U = 3250.5$ $p < .001$), $\tau = 8$ (two tailed, $U = 3307$, $p < .001$) and $\tau = 10$ (two tailed, $U = 3101$, $p < .001$) was observed.

According to the results, our hypothesis H1 was confirmed as multimodal synchronization between the energy of the audio signal of respiration and the kinetic energy of whole body movement allowed us to distinguish between the selected qualities. In particular, audio respiration and kinetic energy were found to be more synchronized in fragmented movements than in fluid movements.

7.5 Classification

We train classifiers per each considered value of τ. We use the same 8 classification algorithms that we have used in the previous Section.

First we compute the following features: the average value ($Avg\,Q^\tau$), the Variance ($Var\,Q^\tau$) and the median value ($Med\,Q^\tau$) of the amount of synchronization Q^τ on the whole segment. The training procedure is the same as in the case of unimodal algorithm (Sect. 6) with 8 fold inner loop and 10 repetitions of the outer loop. The results (see Table 7) are

Table 6 Average values and standard deviations of $Avg\,Q^\tau$, $Var\,Q^\tau$ and $Med\,Q^\tau$ for fluid and fragmented movements

τ	$Avg\,Q^\tau$		$Var\,Q^\tau$		$Med\,Q^\tau$	
	Fluid	Fragmented	Fluid	Fragmented	Fluid	Fragmented
$\tau = 4$	0.130 (0.130)	0.180 (0.151)	0.235 (0.192)	0.295 (0.160)	0.020 (0.139)	0.064 (0.209)
$\tau = 6$	0.184 (0.167)	0.275 (0.185)	0.290 (0.189)	0.351 (0.144)	0.076 (0.262)	0.157 (0.325)
$\tau = 8$	0.249 (0.170)	0.339 (0.189)	0.363 (0.151)	0.386 (0.127)	0.085 (0.277)	0.219 (0.365)
$\tau = 10$	0.284 (0.176)	0.399 (0.215)	0.389 (0.133)	0.401 (0.118)	0.122 (0.322)	0.321 (0.424)

Table 7 Average values of the F-score obtained for 8 algorithms

Algorithm	$\tau = 4$	$\tau = 6$	$\tau = 8$	$\tau = 10$
ADA	0.67 (0.044)	0.68 (0.057)	0.69 (0.032)	0.71 (0.043)
CART	0.65 (0.077)	0.70 (0.066)	0.67 (0.045)	0.70 (0.046)
LDA	0.67 (0.042)	0.67 (0.062)	0.64 (0.036)	0.69 (0.053)
NB	0.65 (0.056)	0.66 (0.051)	0.65 (0.069)	0.72 (0.053)
NN	0.70 (0.040)	0.70 (0.060)	0.68 (0.046)	0.69 (0.074)
RF	0.67 (0.037)	0.63 (0.064)	0.63 (0.035)	0.66 (0.059)
SVM-G	0.67 (0.035)	0.67 (0.070)	0.67 (0.070)	0.64 (0.071)
SVM-LP	0.67 (0.038)	0.66 (0.048)	0.60 (0.040)	0.64 (0.065)

between 0.60 and 0.72 (F-score) depending on the dimension of τ and the classification algorithm.

To check whether there are significant differences between the different machine learning algorithms within the dataset we used ANOVA test. Only significant results are listed below. For the post-hoc tests we used Bonferroni correction.

Given $\tau = 4$, ANOVA did not show significant differences between the classifiers, $F(7, 72) = 1.095$, $p = .096$.

Given $\tau = 6$, ANOVA did not show significant difference between the classifiers, $F(7, 72) = 1.493$, $p = .183$.

Given $\tau = 8$, ANOVA showed significant difference between the classifiers, $F(7, 72) = 3.701$, $p < .005$. Post-hoc tests showed that the F-score for: SVM-LP was significantly lower than the F-score for ADA ($p < .01$) and NN ($p < .05$).

Given $\tau = 10$, ANOVA showed significant difference between the classifiers $F(7, 72) = 2.869$, $p < .05$ while post-hoc tests did not show differences between any specific pair.

7.6 Discussion

In this section, a multimodal approach for the discrimination of fluid and fragmented movements, that is based on Event Synchronization, was presented. First, we observed that there is a significant difference in the amount of the synchronization between fluid and fragmented movements. We used the amount of synchronization as an input to the binary classifier. The highest numerical score was for $\tau = 10$ when NB algorithm (F-score 0.72) was used.

When comparing the results obtained on the same dataset in the Sects. 6.3 and 7.5 it can be seen that the results of second approach are numerically worse. However, it is important to notice that the second solution uses only two very simple features. Even with such a small amount of information as the audio and movement data energy peaks contain, it is possible to compute whether the person moves fluidly or in a fragmented manner.

Our long-term aim is to detect different movement qualities without using a motion capture system. For this purpose, in the future we plan to use the IMU sensors placed on the dancers' limbs, and to estimate their kinetic energy without the need of using motion capture systems (see e.g., [9]).

It is important to notice that we did not ask dancers to play dance patterns, but only to improvise typical movements characterized by the two clusters of movement qualities. All movements were therefore in normal standing positions (e.g. not moving down to the floor). The availability of dancers is to have a better mastery and awareness of movement to obtain a cleaner movement dataset. It is probable that dancers have higher consciousness and control their respiration patterns better than the average people. Thus, the further research is needed to examine if this method can also be successful to analyze average persons, e.g., not dancers.

8 Conclusion

In this paper, we proposed two novel approaches to distinguish fluid and fragmented movements using the audio of respiration. In the first method, MFCC coefficients were extracted from the single exhilarations and used as an input to the binary classification algorithms. The second approach computed the degree of synchronization of multimodal data, consisting of energy peaks of audio signal of respiration and body movements. This degree of synchronization was then used to distinguish between the fluid and fragmented movements. Both methods were validated on the same dataset. According to the results, both techniques were successful to distinguish fragmented and fluid movements and the best results were obtained with SVM-LP (0.87).

The main contributions of this work are:

- according to the authors' knowledge, it is the first attempt to use information extracted from the respiration audio to analyze how a person moves in terms of movement qualities,
- unlike the most of previous works on respiration data, we used a standard microphone placed near to the mouth to capture respiration data,

– whilst most of the works that explored the respiration data mainly focused on the respiration rhythm, however, we investigated other features e.g., intrapersonal synchronization between two modalities.

The paper proves that audio respiration can be useful to recognize how a person moves. While we did not focus on any specific Laban quality but we analyzed very broad movement categories, the positive results obtained in this exploratory study, allows us to assume that, in the future, it will be possible to apply our techniques to recognize more subtle movements qualities from Laban's [27] or other frameworks, e.g., [10]. As the first step in this direction, and inspired by recent works [16] we have been working on creating the multimodal dataset (containing IMU and audio respiration data) of movement qualities of the expressive vocabulary [33]. Furthermore, we expect that the methods proposed in this paper can be useful to detect other human activities, cognitive and emotional states.

Acknowledgements We thank our colleagues at Casa Paganini - InfoMus Paolo Alborno, Corrado Canepa, Paolo Coletta, Simone Ghisio, Ksenia Kolykhalova, Stefano Piana, and Roberto Sagoleo for the fruitful discussions and for their invaluable contributions in the design of the multimodal recordings, and the dancers Roberta Messa, Federica Loredan, and Valeria Puppo for their kind availability to participate in the recordings of our repository of movement qualities.

References

1. Abushakra A, Faezipour M (2013) Acoustic signal classification of breathing movements to virtually aid breath regulation. IEEE J Biomed Health Inf 17(2):493–500. https://doi.org/10.1109/JBHI.2013.2244901
2. Alaoui SF, Caramiaux B, Serrano M, Bevilacqua F (2012) Movement qualities as interaction modality. In: Proceedings of the designing interactive systems conference, DIS '12, pp. 761–769. ACM, New York, NY, USA. https://doi.org/10.1145/2317956.2318071
3. Bateman A, McGregor A, Bull A, Cashman P, Schroter R (2006) Assessment of the timing of respiration during rowing and its relationship to spinal kinematics. Biol Sport 23:353–365
4. Bernasconi P, Kohl J (1993) Analysis of co-ordination between breathing and exercise rhythms in man. J. Physiol 471:693–706
5. Beyan C, Shahid M, Murino V (2018) Investigation of small group social interactions using deep visual activity-based nonverbal features. In: Accepted to ACM multimedia. https://doi.org/10.1145/3240508.3240685
6. Bianchi-Berthouze N (2013) Understanding the role of body movement in player engagement. Taylor & Francis, pp 40–75. https://doi.org/10.1080/07370024.2012.688468
7. Boiten FA, Frijda NH, Wientjes CJ (1994) Emotions and respiratory patterns: review and critical analysis. Int J Psychophysiol 17(2):103–128. https://doi.org/10.1016/0167-8760(94)90027-2
8. Bousmalis K, Mehu M, Pantic M (2009) Spotting agreement and disagreement: a survey of nonverbal audiovisual cues and tools. In: 2009 3rd international conference on affective computing and intelligent interaction and workshops, pp 1–9. https://doi.org/10.1109/ACII.2009.5349477
9. Camurri A, Canepa C, Ferrari N, Mancini M, Niewiadomski R, Piana S, Volpe G, Matos JM, Palacio P, Romero M (2016) A system to support the learning of movement qualities in dance: A case study on dynamic symmetry. In: Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: adjunct, UbiComp '16, pp. 973–976. ACM, New York, NY, USA. https://doi.org/10.1145/2968219.2968261
10. Camurri A, Volpe G, Piana S, Mancini M, Niewiadomski R, Ferrari N, Canepa C (2016) The dancer in the eye: Towards a multi-layered computational framework of qualities in movement. In: 3rd international symposium on movement and computing, MOCO 2016. https://doi.org/10.1145/2948910.2948927
11. Caridakis G, Raouzaiou A, Bevacqua E, Mancini M, Karpouzis K, Malatesta L, Pelachaud C (2007) Virtual agent multimodal mimicry of humans. Lang Resour Eval 41(3):367–388
12. Castellano G, Villalba SD, Camurri A (2007) Recognising human emotions from body movement and gesture dynamics. In: Paiva ACR, Prada R, Picard RW (eds) Affect Comput Intell Interact. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 71–82
13. Cho Y, Julier SJ, Marquardt N, Bianchi-Berthouze N (2017) Robust tracking of respiratory rate in high-dynamic range scenes using mobile thermal imaging. Biomed Opt Express 8(10):4480–4503. https://doi.org/10.1364/BOE.8.004480
14. Codrons E, Bernardi NF, Vandoni M, Bernardi L (2014) Spontaneous group synchronization of movements and respiratory rhythms. PLOS ONE 9(9):1–10. https://doi.org/10.1371/journal.pone.0107538
15. Dempster WT, Gaughran GRL (1967) Properties of body segments based on size and weight. Am J Anat 120(1):33–54. https://doi.org/10.1002/aja.1001200104
16. Fdili Alaoui S, Françoise J, Schiphorst T, Studd K, Bevilacqua F (2017) Seeing, sensing and recognizing laban movement qualities. In: Proceedings of the 2017 CHI conference on human factors in computing systems, CHI '17, pp. 4009–4020. ACM, New York, NY, USA. https://doi.org/10.1145/3025453.3025530
17. Folke M, Cernerud L, Ekström M, Hök B (2003) Critical review of non-invasive respiratory monitoring in medical care. Med Biol Eng Comput 41(4):377–383
18. Forgacs P (1978) Breath sounds. Thorax 33:681–683
19. Glowinski D, Dael N, Camurri A, Volpe G, Mortillaro M, Scherer K (2011) Toward a minimal representation of affective gestures. IEEE Trans Affect Comput 2(2):106–118. https://doi.org/10.1109/T-AFFC.2011.7
20. Hachimura K, Takashina K, Yoshimura M (2005) Analysis and evaluation of dancing movement based on lma. In: Robot and human interactive communication, 2005. ROMAN 2005. IEEE international workshop on, pp. 294–299. IEEE
21. Hoffmann CP, Torregrosa G, Bardy BG (2012) Sound stabilizes locomotor-respiratory coupling and reduces energy cost. PLoS ONE 7(9):e45,206. https://doi.org/10.1371/journal.pone.0045206
22. Huq S, Yadollahi A, Moussavi Z (2007) Breath analysis of respiratory flow using tracheal sounds. In: 2007 IEEE international symposium on signal processing and information technology, pp 414–418. https://doi.org/10.1109/ISSPIT.2007.4458134
23. Jin F, Sattar F, Goh D, Louis IM (2009) An enhanced respiratory rate monitoring method for real tracheal sound recordings. In: Signal processing conference, 2009 17th European, pp 642–645

24. Johansson G (1973) Visual perception of biological motion and a model for its analysis. Percept Psychophys 14(2):201–211
25. Kider J, Pollock K, Safonova A (2011) A data-driven appearance model for human fatigue. https://doi.org/10.2312/SCA/SCA11/119.128
26. Kim J, Andre E (2008) Emotion recognition based on physiological changes in music listening. IEEE Trans Pattern Anal Mach Intell 30(12):2067–2083. https://doi.org/10.1109/TPAMI.2008.26
27. Laban R, Lawrence FC (1947) Effort. Macdonald & Evans, Evans
28. Liu G, Guo Y, Zhu Q, Huang B, Wang L (2011) Estimation of respiration rate from three-dimensional acceleration data based on body sensor network. Telemed J e-Health 17(9):705–711. https://doi.org/10.1089/tmj.2011.0022
29. Niewiadomski R, Chauvigne L, Mancini M, Camurri A (2018) Towards a model of nonverbal leadership in unstructured improvisation task. In: Proceedings of the 5nd international conference on movement and computing, MOCO '18. https://doi.org/10.1145/3212721.3212816
30. Niewiadomski R, Kolykhalova K, Piana S, Alborno P, Volpe G, Camurri A (2017) Analysis of movement quality in full-body physical activities. ACM Trans Interact Intell Syst 9:1. https://doi.org/10.1145/3132369
31. Niewiadomski R, Mancini M, Ding Y, Pelachaud C, Volpe G (2014) Rhythmic body movements of laughter. In: Proceedings of the 16th international conference on multimodal interaction, ICMI '14, pp 299–306. ACM, New York, NY, USA. https://doi.org/10.1145/2663204.2663240
32. Niewiadomski R, Mancini M, Piana S (2013) Human and virtual agent expressive gesture quality analysis and synthesis. In: Rojc M, Campbell N (eds) Coverbal synchrony in human-machine interaction. CRC Press, Boca Raton, https://doi.org/10.1201/b15477-12
33. Niewiadomski R, Mancini M, Piana S, Alborno P, Volpe G, Camurri A (2017) Low-intrusive recognition of expressive movement qualities. In: Proceedings of the 19th ACM international conference on multimodal interaction, ICMI 2017, pp 230–237. ACM, New York, NY, USA. https://doi.org/10.1145/3136755.3136757
34. Oliveira A, Marques A (2014) Respiratory sounds in healthy people: a systematic review. Respir Med 108(4):550–570. https://doi.org/10.1016/j.rmed.2014.01.004
35. Omlin X, Crivelli F, Heinicke L, Zaunseder S, Achermann P, Riener R (2016) Effect of rocking movements on respiration. PLOS ONE 11(3):1–11. https://doi.org/10.1371/journal.pone.0150581
36. Pellegrini R, Ciceri M (2012) Listening to and mimicking respiration: understanding and synchronizing joint actions. Rev Psychol 19(1):17–27
37. Piana S, Alborno P, Niewiadomski R, Mancini M, Volpe G, Camurri A (2016) Movement fluidity analysis based on performance and perception. In: Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems, CHI EA '16, pp 1629–1636. ACM, New York, NY, USA. https://doi.org/10.1145/2851581.2892478
38. Piana S, Coletta P, Ghisio S, Niewiadomski R, Mancini M, Sagoleo R, Volpe G, Camurri A (2016) Towards a multimodal repository of expressive movement qualities in dance. In: 3rd international symposium on movement and computing, MOCO 2016, 5–6 July 2016, Thessaloniki, Greece. https://doi.org/10.1145/2909132.2909262
39. Piana S, Staglianò A, Odone F, Camurri A (2016) Adaptive body gesture representation for automatic emotion recognition. ACM Trans Interact Intell Syst (TiiS) 6(1):6
40. Pollick FE (2004) The features people use to recognize human movement style. In: Camurri A, Volpe G (eds) Gesture-based communication in human-computer interaction. Springer, Berlin, pp 10–19
41. Quian Quiroga R, Kreuz T, Grassberger P (2002) Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. Phys Rev E 66:041,904. https://doi.org/10.1103/PhysRevE.66.041904
42. Ran B, Tal S, Rachelle T, Karen S, Assaf S (2015) Multitask learning for laban movement analysis. In: Proceedings of the 2nd international workshop on movement and computing, MOCO '15, pp 37–44
43. Rao KM, Sudarshan B (2015) A review on different technical specifications of respiratory rate monitors. IJRET Int J Res Eng Technol 4(4):424–429
44. Rehm M (2010) Nonsymbolic gestural interaction for ambient intelligence. In: Aghajan H, Delgado RLC, Augusto JC (eds) Human-centric interfaces for ambient intelligence. Academic Press, Oxford, pp 327–345. https://doi.org/10.1016/B978-0-12-374708-2.00013-9
45. Ruinskiy D, Lavner Y (2007) An effective algorithm for automatic detection and exact demarcation of breath sounds in speech and song signals. IEEE Trans Audio Speech Lang Process 15(3):838–850. https://doi.org/10.1109/TASL.2006.889750
46. Samadani A, Burton S, Gorbet R, Kulic D (2013) Laban effort and shape analysis of affective hand and arm movements. In: 2013 Humaine Association conference on affective computing and intelligent interaction, pp 343–348. https://doi.org/10.1109/ACII.2013.63
47. Schmid M, Conforto S, Bibbo D, DAlessio T (2004) Respiration and postural sway: detection of phase synchronizations and interactions. Hum Mov Sci 23(2):105–119
48. Silang Maranan D, Fdili Alaoui S, Schiphorst T, Pasquier P, Subyen P, Bartram L (2014) Designing for movement: Evaluating computational models using lma effort qualities. In: Proceedings of the 32Nd Annual ACM conference on human factors in computing systems, CHI '14, pp 991–1000. ACM, New York, NY, USA. http://doi.acm.org/10.1145/2556288.2557251
49. Singh A, Piana S, Pollarolo D, Volpe G, Varni G, Tajadura-Jiménez A, Williams AC, Camurri A, Bianchi-Berthouze N (2016) Go-with-the-flow: tracking, analysis and sonification of movement and breathing to build confidence in activity despite chronic pain. Hum Comput Interact 31(3–4):335–383
50. Song I (2015) Diagnosis of pneumonia from sounds collected using low cost cell phones. In: 2015 International joint conference on neural networks (IJCNN), pp 1–8. https://doi.org/10.1109/IJCNN.2015.7280317
51. Sovijarvi AR, Malmberg LP, Paajanen E, Piirila P, Kallio K, Katila T (1996) Averaged and time-gated spectral analysis of respiratory sounds: repeatability of spectral parameters in healthy men and in patients with fibrosing alveolitis. Chest 109(5):1283–1290. https://doi.org/10.1378/chest.109.5.1283
52. Swaminathan D, Thornburg H, Mumford J, Rajko S, James J, Ingalls T, Campana E, Qian G, Sampath P, Peng B (2009) A dynamic bayesian approach to computational laban shape quality analysis. Adv Hum-Comput Interact 362651:17. https://doi.org/10.1155/2009/362651
53. Thaut MH, McIntosh GC, Hoemberg V (2015) Neurobiological foundations of neurologic music therapy: rhythmic entrainment and the motor system. Front Psychol 5:1185. https://doi.org/10.3389/fpsyg.2014.01185
54. Truong A, Boujut H, Zaharia T (2016) Laban descriptors for gesture recognition and emotional analysis. Vis Comput 32(1):83–98
55. Vaessen MJ, Abassi E, Mancini M, Camurri A, de Gelder B (2018) Computational feature analysis of body movements reveals hierarchical brain organization. Cerebral Cortex, [bhy228]. https://doi.org/10.1093/cercor/bhy228
56. Wallbott HG, Scherer KR (1986) Cues and channels in emotion recognition. J Personal Soc Psychol 51(4):690
57. Ward N, Ortiz M, Bernardo F, Tanaka A (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, pp 995–1000. ACM, New York, NY, USA
58. Wlodarczak M, Heldner M (2016) Respiratory belts and whistles: A preliminary study of breathing acoustics for turn-taking. In: Interspeech 2016, pp 510–514. https://doi.org/10.21437/Interspeech.2016-344
59. Yahya O, Faezipour M (2014) Automatic detection and classification of acoustic breathing cycles. In: American society for engineering education (ASEE Zone 1), 2014 zone 1 conference of the, pp 1–5. https://doi.org/10.1109/ASEEZone1.2014.6820648