Evaluating movement quality through intra-personal synchronisation

Nikolas De Giorgis, Enrico Puppo, Paolo Alborno, Antonio Camurri

***Abstract*—We present a method to measure intra-personal synchronisation of movement from motion capture data, and we show that our method is effective in classifying the level of skills of athletes performing karate kata. Our method is based on detecting relevant peaks of acceleration of limbs (arms and legs) and measuring their synchronisation. We run a multi-scale analysis, based on topological persistence, to rank the importance of peaks of acceleration. The resulting impulse signals are processed next with a Multi-Event Class Synchronisation algorithm, in order to define an *Overall synchronisation index* that scores the level of intra-personal synchronisation with a single scalar value. We build a basic multiclass classifier, which uses just the means of indexes computed on the different classes in the training set. We make a statistical analysis and a cross validation of the classifier on real data. Performances by athletes from three levels of skill have been recorded, classified by experts and used to test our method. Cross validation of the classifier is performed by leave-one-out and bootstrap resampling. Results show that our method can classify correctly with very high probability (beyond 99%), while it succeeds on 100% of the data used in cross validation.**

***Index Terms*—Human Movement, Movement Analysis, Movement Qualities, Karate, Event Synchronisation, Intra-personal Synchronisation**

I. Introduction

The analysis of human movement from point-light display or motion capture (*MoCap*) data originated by the seminal work of Johansson [1], and has been investigated for several decades. MoCap data of human movement, although severely impoverished with respect to full video, are still capable to convey high-level perceptual qualities to a human observer, such as sex differences [2] and affect [3], [4]. From a computational perspective, trajectories recorded from the markers in a MoCap sequence are treated as time series and are analysed by applying various types of signal processing techniques [5].

In recent years, higher-level computational analysis techniques emerged, grounded on the results obtained from experimental psychology [6]. Notable examples are the affective bodily expressions in non-verbal communication, and the level of performance in sports and performing arts. Computational models and techniques for the automated analysis of movement in such domains are receiving a growing attention from both the scientific and industrial communities. In the analysis of movement in sport, systems capable to perform automated analysis of movement qualities can be useful for more efficient training [7], and for evaluating the effectiveness of a specific physical gesture in the performance practice.

In this paper, we address the problem of evaluating the overall quality of movement in martial arts performance, starting from MoCap recordings. Figure 1 shows an example of our input data. Our interdisciplinary approach is grounded on computer science, biomechanics, artistic theories [8], and affective computing. More specifically, we aim at quantifying how performances at different levels of expertise are perceived by an external, not necessarily expert, observer. To this aim, we hypothesise that synchronisation of movements between specific limbs can be employed as a main feature to explain the differences between the same *kata* (a movement sequence, or "choreography", in karate) expressed by different athletes with different levels of expertise. The motivations to ground our measure of quality on synchronisation is rooted on well-known assumptions in sport practice and as suggested by karate experts in [9]: an experienced athlete makes a more *neat* execution of a kata with respect to a less skilled athlete, in which the starting and ending moments of each movement (e.g. punches, strikes, kicks) are performed with less or no fluctuations or ripples between joints movements.

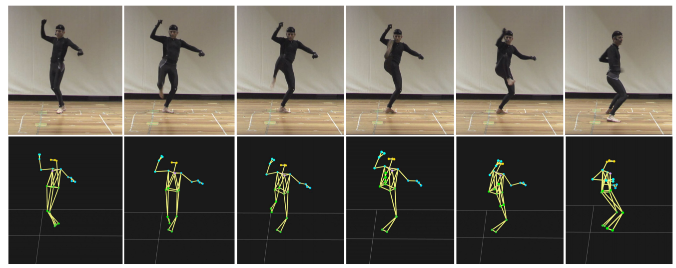


Fig. 1: Video frames and corresponding MoCap data of an athlete performing a kick of the Bassai Dai kata. MoCap data are shown here from the same perspective of video, but they contain full 3D information about markers (coloured joints of the skeleton). We process only the MoCap data.

In music performances, a similar hypothesis has been confirmed by the model of *soft entrainment* [10]: in a joint music performance, a higher level of expressivity corresponds to a higher inter-personal synchronisation at the start and at the end of each musical phrase. In our analogy, we study *intra-personal synchronisation*: a high-level control of the movement corresponds here to a strong synchrony between limbs, at the starting and ending moments of single movements.

We analyse a set of performances (katas) to distinguish and classify them on a measure of *overall quality*. Several problems arise while analysing this kind of data: performing sessions are of different length, due to the different speed at which the athletes perform; there is no external clock to synchronise with, hence, intra-personal synchronisation must be evaluated by relating different parts of the body and by analysing instantaneous events that provide an internal *beat* for synchronisation. This process entails extracting from the MoCap data, assumed as samples of a continuous signal, a set of impulses precisely located in time at relevant instants, which characterise movement. MoCap data are noisy in nature, therefore selecting relevant instantaneous events is all but simple. However, we show that, once the relevant impulse events have been detected, the whole framework becomes surprisingly simple and effective. We tackle these issues by using a multi-scale analysis to filter noise and extract relevant information present in the input signal, which is then fed to an event synchronisation algorithm inspired to [11].

The claim we wish to demonstrate is the high correlation between intra-personal synchronisation and the quality of motion, as well as our ability to evaluate the former in order to classify the latter. To this aim, we deliberately selected a case study that contains neat and sharp movements - i.e., karate *kata* - and a dataset consisting of trials that can be clearly classified by experts into different classes - i.e., no athlete participating in the trials exhibits a border line performance between different classes. We run our algorithms on the dataset of performances, without either introducing any bias, or tweaking our method to achieve the desired results, and we compare our evaluation with respect to benchmark rankings from experts: our experiments confirm that the analysis of synchronisation can be used safely to characterise the quality of performance of athletes from different levels of skill.

II. Related work

Several methods to automatically evaluate the quality of a performance have been proposed and applied in various fields. In dance, the system presented by Alexiadis [12], obtaining data from Kinect, assessed the overall quality by comparing positions and velocity of the various joints with reference performances. Another Kinect-based model, presented by Kitsikidis [13], in order to estimate the performance quality of Greek dance pieces, made use of low-level motion features (e.g., accelerations, velocities, etc.). Muneesawang et al. [14] developed a tool for dance training that, through a visual feedback in the form of a coloured skeleton, mapped the quality and goodness of the movements, i.e., how much they reflect the teachers, to different colours. In rehabilitation, Barrett et al. [15] presented a system to measure the gestures' accuracy through a serious game. Effenberg et al. [16] applied sonification to enhance motor perception and motor control and studied how a real-time auditory feedback applied to arm movement trajectories can improve motor rehabilitation. Ilg et al. [17] developed a model to measure the different levels of ability to perform movement sequences in different sports. In karate, VencesBirto et al. [18] showed a significant difference in kinematics patterns and EMG measurements on the execution of a particular kata (the *choku-zuki* punch), performed by non expert and experienced karatekas. Cynarski et al. [19] measured the differences between novices and experienced karatekas through the analysis of the kinematics of the various joints. Bianco and Tisato [20] proposed an algorithm for karate movement recognition from skeletal motion on a dataset consisting of punches, kicks, and defense karate moves.

The analysis of synchronisation has been studied in many fields. Among investigation of human movements and social interaction, Hwang et al. [21] carried out a study on interpersonal synchronization in which two participants collaboratively control the movement of a spherical virtual object, through a pair of tactile supports connected to a tablet. Demos et al. [22] investigated the effects and influence of differences in the social status of a duet of pianists on their temporal coordination and perceived synchrony. Waterhouse et al. [23] presented a case study on entrainment between a duo of dancers. On the same topic, Lussu et al. [24] analysed synchronisation on multimodal data (respiration energy and body movement energy) to distinguish movements performed with different expressive qualities. Miyake et al. [25] introduced a rehabilitation system based on limbs synchronisation to support and stabilise the walking of patients affected by Parkinson's disease and hemiplegia. Leman et al. [26] concentrated on the effects of beat synchronised walking in human beings on movement timing and vigour. In the specific framework of social interaction studies, Varni et al. [27] investigated how motor synchronisation can be used to analyse social group dynamics and detect dominant members, while Hung and Gatica-Perez [28] used synchronisation to measure the degree of cohesion of the group.

Several methods for the computation of synchronisation have been proposed in literature: correlation analysis techniques have been applied to a wide number of contexts to establish the degree of similarity between two time series [27], [29], [30]. Linear phase correction models describe the process of minimizing the asynchrony by predicting and adjusting the timing of each future movement. Event Synchronization (ES) [11] measures synchronization and time-delay patterns between two time series. The method relies on time difference between events occurrence-timings. Alborno et al. [31] applied ES to demonstrate how to distinguish between full-body movements performed with different expressive qualities (namely rigidity, fluidity, and impulsivity) by applying the Event Synchronisation algorithm on arms and hands kinematics. Wing et al. [32] recently proposed a linear phase correction model to estimate how the members of a string quartet correct asynchronies on tone onsets arising from fluctuations in their individual tempos. An alternative technique, initially applied in signal processing for financial forecasting, is the Wavelet Transform. Wavelets have been increasingly employed in other fields, among these, to the study of human movement. For example, Wong and colleagues [33], applied wavelets to study human movements trajectories. and Fujiwara et al. [34] used them to measure interpersonal synchrony during an unstructured conversation. Finally, Recurrence Quantification Analysis (RQA)[1] has been applied to study and estimate the degree of synchronisation in the study conducted by Varni et al. [35] where RQA is used to investigate entrainment between four violin players measuring the recurrent behaviours of the movement of their heads.

[1] http://www.recurrence-plot.tk/

III. Overview of the method

Our input consists of time series from a motion capture system. Each series contains discrete trajectories from markers placed on the body of the athlete. Our analysis is concentrated on *movements of limbs*. In order to get stable measures, we consider the markers related to each arm and each leg of the participant, forming four independent clusters, which we consider for further processing. Our input, as well as the clustered model, are described in detail in Section IV and summarised in Figure 2.

We aim at analysing such data to estimate the level of *intra-personal synchronisation*, i.e., how synchronously different parts of the body move with respect to each other during performance. Notice that we do not rely on any external source of beat or master clock to synchronise with: as customary in single sport performance, the speed of movement is free.

The lack of an external reference compels us to select instantaneous events originated by the body movement itself, to be used as internal references to evaluate synchronisation. We base our analysis on *peaks of acceleration*. MoCap measures are affected by noise and peaks of acceleration occur very often because of it. In order to get relevant events to synchronise with, we preprocess our data and we score the relevance of each peak through multi-scale analysis based on *topological persistence*. These operations will be described in Section V and are summarised in Figure 3.

Next we take pairs of sequences of events (e.g., left arm vs right arm) and we feed them to our Multi Event Class Synchronisation (MECS) algorithm [36]. We adopted this method due to the fact that, while many of the existing Event Synchronisation algorithms (e.g., [11], [37]) were developed in the context of brain signal analysis, while MECS was created with the purpose of studying multimodal human-human and human-machine interaction. We are interested in measuring the overall degree of motor synchronisation considering an high number of measures (the various joints' acceleration) in which we detect relevant peaks; the chosen algorithm provides a way to compute a single synchronisation index as a function of the events occurring in N time series. We match corresponding events and we weight their amount of synchronisation with a linear kernel that is a decreasing function of the time shift separating them. These measures provide an event-wise estimate of synchronisation; we integrate such measures on the whole sequence to obtain a global estimate in the form of an Overall Synchronisation Index. More details about the synchronisation algorithm can be found in Section VI.

We propose the degree of synchronisation as a measure of quality of a performance, and this criterion is tested against ground truth in Section VII. We also build a basic classifier that makes use of just the means of the Overall Synchronisation Indexes measured on a training set. The classifier, as well as its cross-validation, are discussed in Section VII-C.

IV. Recording Dataset

We have used the dataset presented in [9]. Recordings were acquired using a motion capture system Qualysis with 9 high resolution cameras and a 250Hz frame rate. Post-processing has been applied to get a labeled dataset, with a skeleton model, in which noise from "ghost" or jitter markers has been as much as possible cleaned out.

#	Description	Years of practice	Experience (from 1-low to 5-high)
1	Adult, Male, Karate teacher	> 15	5
2	Adult, Male, Black Belt	> 10	4
3	Adult, Female, Black Belt	> 10	4
4	Teenager, Male	> 5	3
5	Adult, Male participated in the World Championships	> 15	5
6	Teenager, Male	> 5	3
7	Teenager, Male	> 5	3

TABLE I: Subdivision of participants

A total of 7 athletes participated in the recordings; they were chosen with different skills and levels to have a various representation of abilities of execution; their ability has been assessed by experts on a conventional scale from 1 to 5, while only levels from 3 to 5 are represented in the dataset. The subdivision is summarized in table I.

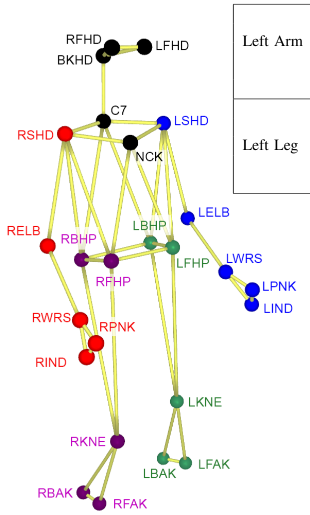
The participants performed two different katas, namely *Heian Yondan* and *Bassai Dai*. Each athlete performed 2 or 3 trials of each kata. The final datasets consists of 32 recordings, divided as follows:

- *Level 3 - Junior, Brown belt*: 7 recordings of the Heian Yondan kata and 7 of the Bassai Dai
- *Level 4 - Senior, Black belt 1st dan*: 4 for Heian Yondan and 5 for Bassai Dai
- *Level 5 - Master, beyond 4th dan*: 5 for Heian Yondan and 4 for Bassai Dai

Athletes were let free to perform at their own speed and rhythm with no interventions, even in post processing, to match the lengths of trials or align the recordings. As a result, recordings have differences both in length (*Bassai Dai* ranges from 71 to 115 seconds, *Heian Yondan* from 50 to 97 seconds), and also in speed of relative body parts.

Motion tracking has been obtained by 25 markers placed on the body (as shown in Figure 2), each providing a 3D trajectory of samples at 250hz. Every sample consists of a triple (x, y, z) of coordinates representing the position of a marker at each time instant, in a calibrated reference system.

Since our analysis is focused on the movements of the limbs, we do not use every marker of the model but we rather extract four clusters (one for each limb, see Figure 2) in order to exploit redundant information and obtain a reduced yet more stable representation. Five markers on the head and upper torso have been set out of our analysis, because they do not have any symmetric counterpart in the body; moreover, movements of the head in karate often must precede the movement of limbs, thus resulting not synchronised with them in a correct performance. This particular clusterisation is relevant in karate analysis, but it is not unique: the particular choice in another context (e.g. dance) would be dependant on the body parts that are expected to be synchronised.



Left Arm	LIND	Right Arm	RIND
	LPNK		RPNK
	LELB		RELB
	LWRS		RWRS
	LSHD		RSHD
Left Leg	LFHP	Right Leg	RFHP
	LBHP		RBHP
	LKNE		RKNE
	LBAK		RBAK
	LFAK		RFAK

Fig. 2: The MoCap skeleton. In red, green, purple and blue the groups of markers used to define the four clusters, summarised in the side table. L and R at the beginning discrimate between left and right. The other three letters indicate the marker position as follows: *IND*=index finger, *PNK*=pinkie finger, *ELB*=elbow, *WRS*=wrist, *SHD*=shoulder, *FHP* and *BHP*=front and back of the hip, *KNE*=knee, *FAK* and *BAK*=front and back of the ankle.

V. Extracting instantaneous events

As outlined before, we aim at analysing the level of synchronisation between limbs. The feature we found to be more representative for our analysis is given by peaks of limbs' acceleration (and deceleration) that allow us to distinguish the initial and final phases of the basic movements, such as punches, strikes, kicks, steps, parry and block actions. Since recordings are quite noisy, in order to perform a stabler analysis, we extract a smoothed version of this feature: we compute the velocity of the clusters' barycenters at each frame, then we apply a standard Savitzky-Golay filter [38] to this signal. Acceleration is derived next from the smoothed velocity. Examples of a raw velocity signal, its corresponding smoothed signal and the derived acceleration are provided in Figure 3, by the magenta (a), green (b) and red (c) graphs, respectively.

Peaks of acceleration and deceleration characterise instants of time that are relevant for our analysis, but they also appear along trajectories, because of noise, uncertainty and ripple in the movement. Since data are rather noisy, isolating relevant peaks from unimportant ones is a challenging task. We tackle this problem by performing a topological analysis, which gives us a multi-scale representation of our signal in terms of its critical points' *persistence*, i.e. a measure of importance related to their difference in amplitude. Persistence induces a total ranking of critical points, which can be used to discriminate noisy and spurious peaks from relevant ones by simple thresholding.

Topological persistence is a concept related to Morse theory [39], which provides a characterisation of a function in terms of topology of its level sets. It can be seen as a filtering process which takes place in the *amplitude* domain, i.e. by progressively removing pairs of critical points depending on their relative values. In a one-dimensional signal, persistence is obtained through a flooding process of *basins,* each expanding from a local minimum: each time a basin gets filled at one of its sides (i.e., along the path connecting its local minimum and the lowest of its two adjacent maxima), this basin gets merged with one of its two neighbours; contextually, the minimum corresponding to the basin and the maximum that has been flooded are removed, and the persistence value of both of them is set at their difference in amplitude. For a graphical depiction of the process, see Figure 4. Notice that the merge of basins and the corresponding deletion of flooded pairs of critical points changes the adjacencies of basins: a minimum and a maximum that were relatively far in the original sequence become adjacent during the process, when all critical points between them have been filtered out.

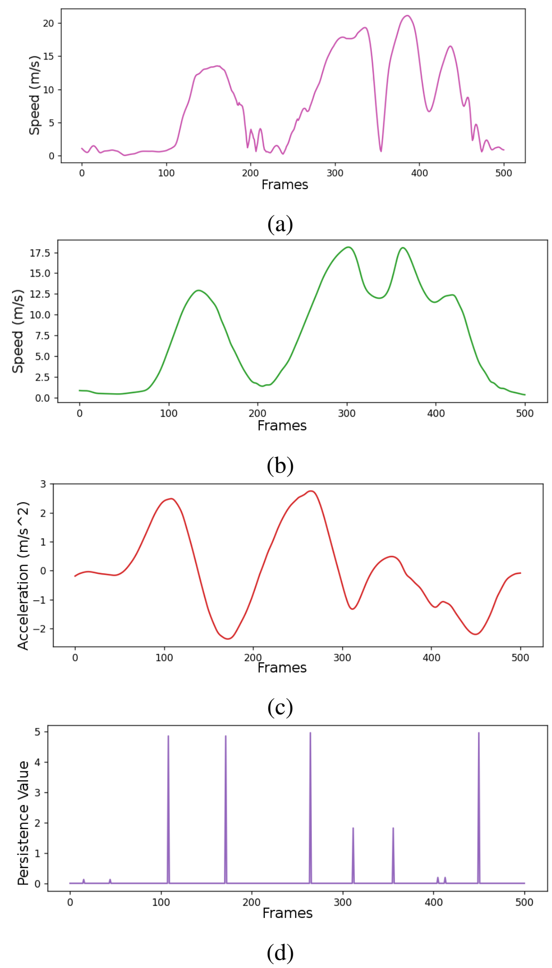


Fig. 3: Detection of instantaneous events on a single trajectory. From the top: velocity from raw data (a); smoothed velocity (b); acceleration from smoothed velocity (c); Persistency Index (d). Time shift related to smoothing does not harm synchronisation, since it has the same effect on all signals.

More formally, let us consider a piecewise-linear function f defined discretely by its value at sample points regularly spaced in the integer domain and extended to $\mathbb{R}$ by linear interpolation, and the set of its critical points. A point i is

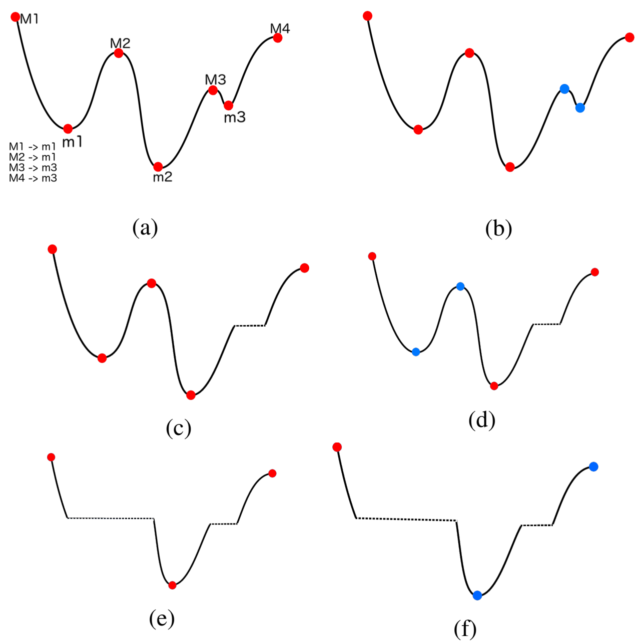


Fig. 4: Steps of persistence computation: (a) An input signal f with its critical points highlighted in red, and an initial pairing. (b) The first step of the filtering: a pair of critical points that is about to be removed is highlighted in blue. (c) An equivalent signal after removing the critical points: the basin located at the removed minimum has been merged into the one to its left. (d) Another pair of critical points to be removed is marked in blue. (e) An equivalent signal after removing the critical points: just one basin has been left. (f) The last pair of critical points to be removed is marked in blue. Dotted lines are just placeholders to denote that in such interval the function is considered *as if* it were monotonic. Notice that the function is *not* modified by the algorithm, only the relative adjacency of critical points changes.

called

- a *maximum* (M) if $f(i) > f(i-1)$ and $f(i) > f(i+1)$
- a *minimum* (m) if $f(i) < f(i-1)$ and $f(i) < f(i+1)$

The algorithm computes persistence as follows:

1) Initially, each maximum M_i is paired to its neighbouring minimum m_j, whose difference in value from M_i is minimal. Figure 4 (a) shows an example of the initial pairing. We take a bookkeeping of the current pairings in a map data structure, as they may become obsolete during processing (see point (3), last step).
2) Then, all maxima are stored in a priority queue Q, where priority is set by the difference in value between M_i and the minimum paired with it: maxima with lower difference have higher priority.
3) Maxima are progressively popped from queue Q. Each time a maximum M_i is popped: if its pairing has become obsolete, then it is discarded; otherwise, the absolute difference between M_i and its paired minimum m_j is written in output as their persistence, and the pairings of their neighbouring critical points are updated as follows:
   - M_i and m_j are excluded and marked as *inactive*; from a theoretical point of view, this is like removing the pair of critical points selected – e.g. the two blue points in Figures 4 (b), (d), and (f) – thus obtaining a new function with two less extrema (see Figure 4 (c), (e)). From the perspective of the flooding process, it means that the basin with the local minimum in m_j is merged into the adjacent one on the other side of M_i.
   - The pairing of the closest active maximum M_k on the other side of M_i with respect to m_j might need to change (in case it were also paired with m_i): in this case, the new pairing of M_k is determined by looking at its active neighbuor minima and choosing the one with the closest value to it. The bookkeeping of pairings is updated by assigning to M_k a new paired minimum, and M_k is then reinserted into Q, thus making any older instance of itself in the queue obsolete; since the auxiliary data structure is always up-to-date, an instance of a critical point M_k in Q is considered obsolete if and only if its paired minimum m is different from the one in the bookkeeping.

   The process ends when there are no more elements in the priority queue. The last maximum remaining in the filtered signal is assigned an arbitrarily high value of persistence, larger than the maximum persistence computed previously.

At the end of the process, each critical point c_k of the input signal has been assigned a real number p_k, which is the computed persistence for that critical point. The result of the analysis is encoded into a time series P, defined as follows:

$$Pi = \begin{cases} p_k & \text{if } f(i) \text{ is a critical point } c_k \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Therefore, the output of our analysis is a sequence of impulses, which we call a *Persistence Index* (PI for short). The purple graph in Figure 3(d) depicts the Persistence Index corresponding to the acceleration in the red graph. The Persistence Index is taken in input by our synchronisation algorithm.

VI. Multi Event Class Synchronisation

The output of persistence analysis consists of the following time-series:

$$ts_f \quad \text{with} \quad f \in \{leftArm,\ rightArm,\ leftLeg,\ rightLeg\}$$

containing the values of the Persistence Index (or *events*) at each frame.
We now consider pairs of time series ts_f and ts_g, related to different limbs, and we measure their mutual synchronisation. Our method relies on two parameters:

1) A threshold ρ to select only a subset of events from ***PI*** (considered as the relevant events);
2) A maximum time lag τ between synchronous events.

We first threshold both time series: only events with $PI > \rho$ are maintained. Then, we apply an extended variant of the

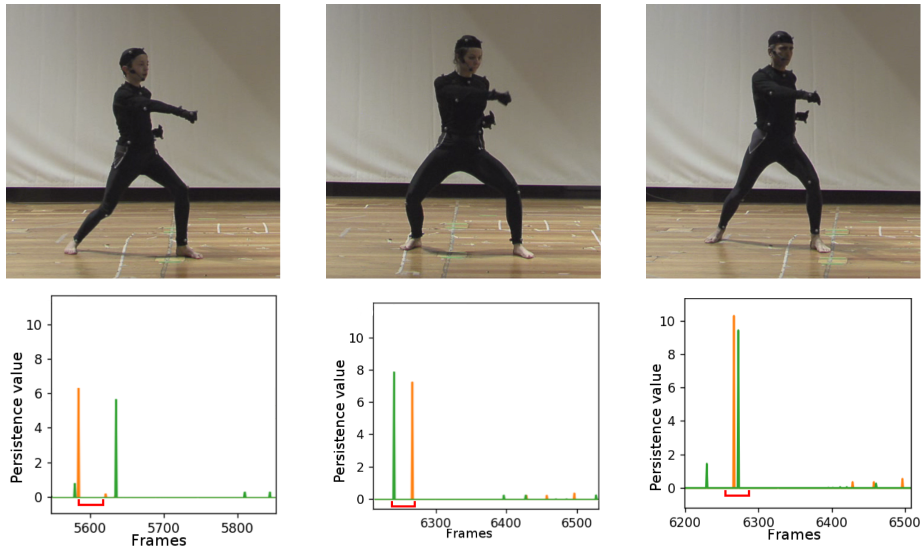


Fig. 5: The same action - the ending of a punching phase - performed by athletes at different levels of skill. From left to right: Level 3, 4 and 5 (top); and their corresponding PI's (bottom). In orange, PI of the right arm; in green, PI of the left arm. The red bracket is the size of the synchronisation kernel $\tau = 40$ used for our analysis, showing that movements of the Level 3 performer will not count as synced, while both Level 4 and Level 5 will contribute positively to the synchronisation index, but with different intensities. Only the two high peaks are considered, while the remaining peaks are below the threshold of relevance ρ. This action occurs at different frames in the different trials.

Event Synchronisation algorithm [11]: the Multi Event Class Synchronisation (MECS) algorithm [36]. Similarly to [11], MECS consists of two steps: the algorithm first detects the coincident events in different time series (*coincidence detection*) and counts them; then the number of coincidences is normalized with respect to the total number of possible coincidences that may happen (*normalization*). Besides, MECS modulates the contribution of each pair of coincident events as a decreasing function of their difference in time.

Let t_f^A with $A = 1, ..., m_f$ and t_g^B with $B = 1, ..., m_g$ be the times of events A and B occurring in sequences ts_f and ts_g, respectively, where m_f and m_g represent the total number of events in the time-series ts_i and ts_j, respectively. A pair of events A and B contribute to the synchronisation index if and only if they occur within a time interval (coincidence window) not larger than τ.

Coincidences are detected by a simple parallel scan of the two time series, like in a merge algorithm for sorted sequences: pairs of events A and B that are consecutive in the merged sequence and such that t_i^A and t_j^B differ for no more than τ are paired, while events that have no neighbour within a time distance of τ are skipped; note that each event A can be paired with just one event B in the other sequence, and vice-versa.

For each pair of coincident events A and B with time occurrences t_f^A and t_g^B we set a synchronisation rate in the interval $[0, 1]$:

$$c_{f,g}(A, B) = \psi_\tau(d(A, B))$$

where d measures the temporal distance between events:

$$d(A, B) = |t_i^A - t_j^B|$$

and ψ_τ is a kernel depending on parameter τ, i.e., a decreasing function with finite support in interval $[0, \tau]$. Several kernels can be used, e.g., with exponential or sigmoid decay; we found a linear ramp to give the best results in our case:

$$\psi_\tau(t) = \begin{cases} 1 - \frac{t}{\tau} & \text{if} \quad 0 \leq t \leq \tau \\ 0 & \text{otherwise} \end{cases}$$

In our experiments, we analysed the impact of varying τ on the results obtained, and we show the findings in Section VII-A. Figure 5 shows a comparison of the synchronisation measured on the arms of athletes at different levels while executing the same action.

Similarly to [11], the *average degree of synchronisation* Q_τ for a pair of time-series ts_i and ts_j is finally given by:

$$Q_\tau = \frac{\sum c_{f,g}(A, B)}{(m_f + m_g)/2} \tag{2}$$

Note that Q_τ takes into account not only the number of synchronised events with respect to the number of occurring events, as in [11], but it also provides a measure of the synchronisation strength in time, due to our weighted version of the synchronisation index $c_{i,j}$. The average degree of synchronisation associates a unique number in the interval $[0, 1]$ to each time series, which will be used in our experiments to rank the overall quality of performance.

VII. Data Analysis

We processed all data described in Section IV without introducing any bias and by setting the same parameters for all of them. For all recordings, we computed the four PI series of clusters corresponding to arms and legs (see Section V). Pairs

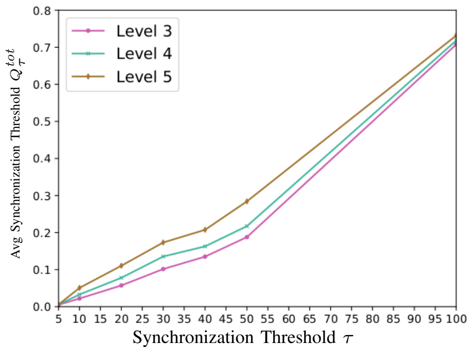


Fig. 6: Average of the overall synchronisation index Q_τ^{tot} computed on the trials of each level at different values of the synchronisation threshold τ (the number of frames).

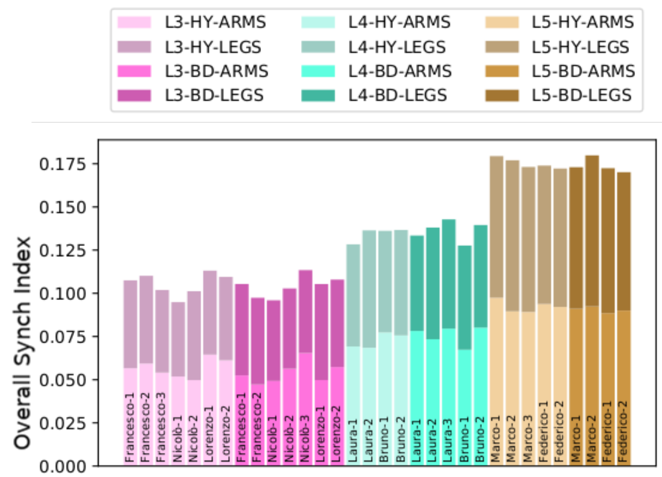


Fig. 7: The overall synchronisation index for all trials in the dataset. Each bar represents the value of Q_τ^{tot} and is divided into two segments, representing Q_τ^{arms} (lower, lighter) and Q_τ^{legs} (upper, darker), respecitvely. Different hues correspond to the three levels (L3 magenta, L4 cyan, L5 brown). Less and more saturated colours correspond to the two different katas (Heian Yondan lighter; Bassai Dai bolder). Labels inside bars permit to identify different takes by the same subject.

of PI series were used next as input to the MECS algorithm. The result for each pair of series is a value Q_τ, in the range $[0, 1]$, where 0 means a total lack of synchronisation, while 1 means that all the detected events of the two limbs are perfectly synchronous (see Section VI).

Among all six possible pairs of PI series, we concentrate our analysis on two pairs: left arm vs right arm; and left leg vs right leg. Mixed combinations of an arm vs a leg were also tried, but provided less significant results; this is not surprising because many movements in karate involve coordination of arms while legs remain static or have a role of maintaining postural stability. In mixed arm-leg combinations, a large number of events in one sequence find no mate in the other, hence giving small values of Q_τ in all cases.

In the end, for each trial we obtain two values Q_τ^{arms} and Q_τ^{legs} to measure the level of synchronisation between arms and between legs, respectively. Since such indexes are on a congruent scale, we define the *Overall synchronisation index*

$$Q_\tau^{tot} = Q_\tau^{arms} + Q_\tau^{legs} \tag{3}$$

which will be used in our statistical tests. The purpose of our analysis is to show that index Q_τ^{tot} provides sufficient information to discriminate the level of performance among the three levels of athletes, congruent with annotations by experts (see Section IV). In the following, we first show how our index characterises the three different groups, then we test the discriminative power of a simple classifier based on it.

A. Parameters setting

As described in Section VI, our method uses two parameters ρ and τ. Persistency Indexes of all series were normalised on interval $[0, 1]$ and then thresholded with a value $\rho = 0.15$. This means that a peak of acceleration is considered relevant if its persistence is larger than 15% of the dynamic range of accelerations. This threshold was found empirically to preserve the relevant peaks, while discarding the residual noise.

Parameter τ determines the size of the kernel used to weight synchronisation. The value of Q_τ, hence Q_τ^{tot}, for the same data is monotonically increasing with τ. However, if τ is too small, too many pairs of potentially synchronous events are missed; while if τ is too large, the value of Q_τ^{tot} becomes less discriminative. For all trials, we carried out our analysis with multiple values of τ, to estimate the impact of this parameter on results. As summarised in the chart of Figure 6, the choice of τ is not critical: performances from athletes of higher level of skill return consistently higher values of Q_τ^{tot}; while such values remain discriminative on a reasonably large range of values of τ. From the chart, it appears reasonable to place τ in the range from 10 to 50 frames, i.e., from 1/25 to 1/5 of a second. In what follows, we set $\tau = 40$, i.e., about 1/6 of a second.

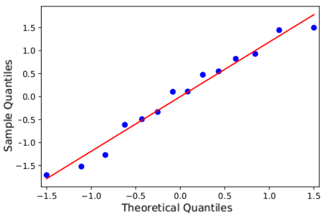
B. Statistical analysis

The barchart in Figure 7 shows the overall synchronisation indexes for all the trials, arranged in the three groups corresponding to the different levels of skill. Note that all scores for the group at Level 3 (magenta, left) are smaller than the scores of group at Level 4 (cyan, center), which in turn are smaller of scores at Level 5 (brown, right). The visual analysis of this chart suggests that the overall synchronisation index succeeds in discriminating between different levels of skill. In order to provide a more rigorous evaluation of the discriminative power of our index, we have employed standard tools from statistical analysis.

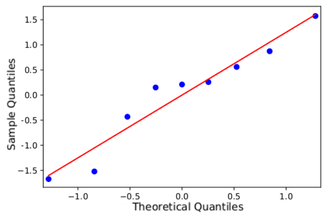
We first run a test aimed at discarding the null hypothesis, i.e., the assumption that our index characterises all data as coming from the same population. Due to the small sample size, we have computed the *effect size* using *Cohen's d* [40]. The effect size can be seen as a quantitative measure of the difference between two populations. To obtain the value for two classes of size n_1 and n_2 with respectively average values of M_1 and M_2 and standard deviations SD_1 and SD_2, we first compute the *Pooled Standard Deviation:*

Classes compared	*Cohen's d*
Level 3 vs Level 4	3.477
Level 4 vs Level 5	4.277
Level 3 vs Level 5	6.523

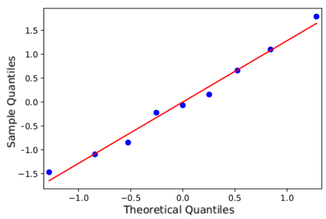
TABLE II: Effect size between different classes computed using Cohen's d.



(a) Level 3



(b) Level 4



(c) Level 5

Fig. 8: Q-Q plots for the three groups at Levels 3 (a), 4 (b), and 5 (c) support the null hypothesis for each group.

$$SD_{pooled} = \sqrt{\frac{|(n_1 - 1) * SD_1^2| + |(n_2 - 1) * SD_2^2|}{n_1 + n_2 - 2}} \tag{4}$$

and we then obtain *Cohen's d* as:

$$d = \frac{M_2 - M_1}{SD_{pooled}} \tag{5}$$

As initially suggested by Cohen [41] and expanded by Sawilowsky [42], a value of $d = 0.8$ means a large effect size, and values of $d = 1.2$ and $d = 2$ represent respectively *very large* and *huge* effect size. For our computed overall synchronisation indices, we get the results summarised in table II, which are all much larger than 2, thus supporting our hypothesis of significant difference between the populations, while also implying a high statistical power of the test despite the low sample size [43]. These results were also corroborated by running ANOVA 1-way test and Tukey HSD Test, with similar results to support our hypothesis; we do not report them here for brevity.

Next, we verify that it is reasonable to assume samples from each single group to come from a normal distribution (null hypothesis for each group). This allows us to model each group by its distribution, in order to study the amount of overlap between groups. Since our samples are generated by taking several takes of two different katas from several different athletes, the null hypothesis is all but a foregone conclusion. We have ran a Shapiro-Wilk test [44], which returns p-values of 0.635, 0.504, and 0.943 for Levels 3, 4 and 5, respectively.

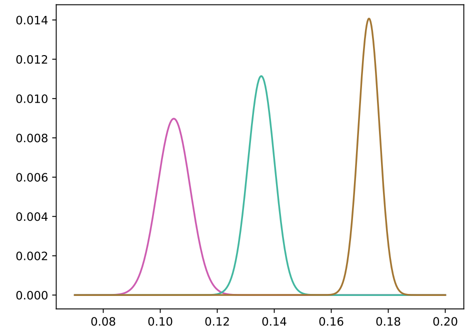


Fig. 9: The Gaussians representing the ideal normal distributions of the three groups. Level 3 is represented in magenta (left), level 4 in cyan (center) and level 5 in brown (right). The area of overlap below different curves gives the probability that an element is misclassified. Overlap between Level 3 and Level 4 is tiny; while overlap between Level 5 and the other two groups is nearly null.

Such large values allow us to say there is no evidence that the null hypothesis can be discarded, and suggest that each group may reasonably come from a normal distribution, but data are too few to support a stronger claim. However, also the Q-Q Plots depicted in Figure 8 show evidence that data are well fit by straight lines in all three cases, thus supporting the assumption that each group comes from a normal distribution. Therefore, we proceed under this hypothesis.

We have computed the mean and variance of each group to compare their distributions visually and numerically. Figure 9 shows the three Gaussians corresponding to such means and variances, which are clearly well distinct. We have computed the area of overlap of such Gaussians by using [45], in order to estimate the probability that an element is misclassified:

- The Gaussian of Level 3 overlaps the other two for just 0.0032, hence an element of Level 3 is classified correctly with probability 0.9968
- The Gaussian of Level 4 overlaps the other two for just 0.0032, hence an element of Level 3 is classified correctly with probability 0.9968
- The overlap of the Gaussian of Level 5 with the other two is below the precision of the method of computation that we have used; hence an element of Level 5 is classified correctly with probability $1 - \varepsilon$ where ε represents the numerical accuracy of the method.

C. A basic classifier

On the basis of previous analysis, we define a basic classifier as follows: let μ_3, μ_4 and μ_5 be the means of the Q_τ^{tot} indexes for the three groups of observed trials, respectively. Given a new trial, let q be its Q_τ^{tot} index. We classify the new trial to belong to:

- Level 3 if $q \leq (\mu_3 + \mu_4)/2$;
- Level 4 if $q > (\mu_3 + \mu_4)/2$ and $q \leq (\mu_4 + \mu_5)/2$;
- Level 5 if $q > (\mu_4 + \mu_5)/2$.

Better thresholds can be obtained by computing the intersection points of Gaussians of consecutive groups (which is easily done with [45]); however, our data are so well separated that we did not try a setting finer than the midpoint between consecutive means.

We test the performance of our classifier by cross validation through leave-one-out and bootstrap resampling. The leave-one-out technique culls one of the samples, say s, builds the classifier on the remaining samples and tries to classify s through it. We repeated the test by culling in turn all data and we obtained 100% of correct classifications.

The bootstrap resampling builds three new groups by random sampling from each group, with replacement, the same number of elements of the original group. Since resampling is made with replacement, some data are sampled multiple times, while some data remain out-of-bag. The classifier is built on the resampled groups, and it is tested on the out-of-bag data. We repeated the bootstrap resampling 100,000 times and we tested the resulting classifiers on a total of over 1 million data (out-of-bag). Also in this case, we obtained 100% of correct classifications.

VIII. Conclusion and Future work

We have presented a method to estimate movement quality in karate by studying how much the limbs are synchronised during relevant motion phases. Our approach demonstrates to be extremely robust on real examples consisting of MoCap data from 32 performances by athletes from three different levels of skill, as classified by experts in this martial art. A basic classifier built on our analysis succeeds in 100% of the subjects according to standard tests of cross validation, suggesting strong correlation between the level of skill of an athlete and her intra-personal synchronization.

Because of the lack of data, we could not test our approach on more than three classes, but we believe that there is good potential to apply the same approach also to several classes and to larger datasets in which the separation between consecutive classes is less sharp. Of course, we expect our method to decrease its performances in matching human classification on datasets that contain border line cases, for which even classification by different experts may disagree. It should be noted that our analysis is completely independent from any assumptions on the data: all the recordings are treated in the same way, independently of length, speed and specific performance. We neither needed to tweak our method, nor to bias data in any way in order to achieve the desired classification. This makes our technique easily applicable with little effort to other scenarios. One possible application is to measure the level of synchronisation of movements of dancers.

Events detected with our multi-scale approach could be used to derive also other measures of quality. For instance, kata usually contain many sudden and fast movements, and intuition suggests that experienced athletes are better at having a cleaner transition when starting the movement and, most important, when ending it. We plan to extend our analysis to measure the level of "cleanness" of motion and use it as another way to discriminate between different levels of skill. We are also working on the application of the same analysis to the automated segmentation of different movements (e.g dance movements): relevant peaks of acceleration, as detected and ranked by the multi-scale analysis, give us a robust way to find the beginning and end of each relevant movement; also, the ranking among extracted peaks allows us to tune the granularity of such segmentation.

Acknowledgments

This research has received funding from the European Union Horizon 2020 research and innovation programme under grant agreements No 645553 (DANCE) and No 688865 (WhoLoDanceE). Support from MIUR PRIN 2015 under grant agreement No 2015B8TRFM (DSurf) is also acknowledged. Federico Campofiorito, Sensei Marco Soggiu and his students are kindly acknowledged for providing their performances for the trials.

References

[1] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception & psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.

[2] L. T. Kozlowski and J. E. Cutting, "Recognizing the sex of a walker from a dynamic point-light display," *Attention, Perception, & Psychophysics*, vol. 21, no. 6, pp. 575–580, 1977.

[3] F. E. Pollick, H. M. Paterson, A. Bruderlin, and A. J. Sanford, "Perceiving affect from arm movement," *Cognition*, vol. 82, no. 2, pp. B51–B61, 2001.

[4] W. H. Dittrich, T. Troscianko, S. E. Lea, and D. Morgan, "Perception of emotion from dynamic point-light displays represented in dance," *Perception*, vol. 25, no. 6, pp. 727–738, 1996.

[5] A. Bruderlin and L. Williams, "Motion signal processing," in *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '95*. New York, NY, USA: ACM, 1995, pp. 97–104.

[6] B. L. Q. University. (2018) Bml walker. [Online]. Available: https://www.biomotionlab.ca/Demos/BMLwalker.html

[7] A. O. Effenberg, U. Fehse, G. Schmitz, B. Krueger, and H. Mechling, "Movement sonification: effects on motor learning beyond rhythmic adjustments," *Frontiers in Neuroscience*, vol. 10, p. 219, 2016.

[8] R. Laban and F. C. Lawrence, *Effort*. Macdonald & Evans, 1947.

[9] K. Kolykhalova, A. Camurri, G. Volpe, M. Sanguineti, E. Puppo, and R. Niewiadomski, "A multimodal dataset for the analysis of movement qualities in karate martial art," in *Intelligent Technologies for Interactive Entertainment (INTETAIN), 2015 7th International Conference on*. IEEE, 2015, pp. 74–78.

[10] T. Yoshida, S. Takeda, and S. Yamamoto, "The application of entrainment to musical ensembles," in *Proceedings of the Second International Conference on Music and Artificial Intelligence (ICMAI)*, 2002.

[11] R. Quian Quiroga, T. Kreuz, and P. Grassberger, "Event synchronization: a simple and fast method to measure synchronicity and time delay patterns," *Physical Review E*, vol. 66, no. 4, 2002.

[12] D. S. Alexiadis, P. Kelly, P. Daras, N. E. O'Connor, T. Boubekeur, and M. Moussa, "Evaluating a dancers performance using kinect-based skeleton tracking," in *Proceedings of the 19th ACM International Conference on Multimedia - MM 11*. New York, NY, USA: ACM, 2011, pp. 659–662.

[13] A. Kitsikidis, K. Dimitropoulos, E. Yilmaz, S. Douka, and N. Grammalidis, "Multi-sensor technology and fuzzy logic for dancer's motion analysis and performance evaluation within a 3d virtual environment," in *Proceedings of 8th International Conference on Universal Access in Human-Computer Interaction*. Springer, 2014, pp. 379–390.

[14] P. Muneesawang, N. M. Khan, M. Kyan, R. B. Elder, N. Dong, G. Sun, H. Li, L. Zhong, and L. Guan, "A machine intelligence approach to virtual ballet training," *IEEE MultiMedia*, vol. 22, no. 4, pp. 80–92, Oct 2015.

[15] C. Barrett, J. Brown, J. Hartford, M. Hoerter, A. Kennedy, R. Hassan, and D. Whittinghill, "Estimating gesture accuracy in motion-based health games," *Journal of Virtual Reality and Broadcasting*, vol. 11(2014), no. 8, 2014.

[16] G. Schmitz, J. Bergmann, A. O. Effenberg, C. Krewer, T.-H. Hwang, and F. Müller, "Movement sonification in stroke rehabilitation," *Frontiers in Neurology*, vol. 9, 2018.

[17] W. Ilg, J. Mezger, and M. Giese, "Estimation of skill levels in sports based on hierarchical spatio-temporal correspondences," in *DAGM03*, 2003, pp. 523–531.

[18] A. M. VencesBrito, M. A. R. Ferreira, N. Cortes, O. Fernandes, and P. Pezarat-Correia, "Kinematic and electromyographic analyses of a karate punch," *Journal of Electromyography and Kinesiology*, vol. 21, no. 6, pp. 1023–1029, 2011.

[19] W. J. Cynarski, W. Wajs, and A. Vences-Brito, "Improving the movements of basic karate techniques with the use of motion capture and mathematical modeling. the concept of a research project," *Ido Movement for Culture. Journal of Martial Arts Anthropology*, vol. 14, no. 1, pp. 39 – 53, 2014.

[20] S. Bianco and F. Tisato, "Karate moves recognition from skeletal motion," *IS&T/SPIE Electronic Imaging*, vol. 8650, pp. 86 500K1–86 500K10, 2013.

[21] T.-H. Hwang, G. Schmitz, K. Klemmt, L. Brinkop, S. Ghai, M. Stoica, A. Maye, H. Blume, and A. O. Effenberg, "Effect- and performance-based auditory feedback on interpersonal coordination," *Frontiers in Psychology*, vol. 9, p. 404, 2018.

[22] A. P. Demos, D. J. Carter, M. M. Wanderley, and C. Palmer, "The unresponsive partner: roles of social status, auditory feedback, and animacy in coordination of joint music performance," *Frontiers in Psychology*, vol. 8, p. 149, 2017.

[23] E. Waterhouse, R. Watts, and B. E. Bläsing, "Doing duo–a case study of entrainment in william forsythes choreography duo," *Frontiers in Human Neuroscience*, vol. 8, p. 812, 2014.

[24] V. Lussu, R. Niewiadomski, G. Volpe, and A. Camurri, "Using the audio respiration signal for multimodal discrimination of expressive movement qualities," in *7th Int. Workshop on Human Behavior Understanding (HBU'2016) in conjunction with ACM Multimedia*. ACM, 2016.

[25] Y. Miyake, "Interpersonal synchronization of body motion and the walk-mate walking support robot," *Robotics, IEEE Transactions on*, vol. 25, no. 3, pp. 638–644, 2009.

[26] M. Leman, D. Moelants, M. Varewyck, F. Styns, L. van Noorden, and J.-P. Martens, "Activating and relaxing music entrains the speed of beat synchronized walking," *PloS one*, vol. 8, no. 7, p. e67932, 2013.

[27] G. Varni, G. Volpe, and B. Mazzarino, "Towards a social retrieval of music content." in *SocialCom/PASSAT*, 2011, pp. 1466–1473.

[28] H. Hung and D. Gatica-Perez, "Estimating cohesion in small groups using audio-visual nonverbal behavior," *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 563–575, Oct 2010.

[29] E. Ono, M. Motohasi, Y. Inoue, D. Ikari, and Y. Miyake, "Relation between synchronization of head movements and degree of understanding on interpersonal communication," in *System Integration (SII), 2012 IEEE/SICE International Symposium on*. IEEE, 2012, pp. 912–915.

[30] K. Yun, K. Watanabe, and S. Shimojo, "Interpersonal body and neural synchronization as a marker of implicit social interaction," *Scientific reports*, vol. 2, p. 959, 2012.

[31] P. Alborno, S. Piana, M. Mancini, R. Niewiadomski, G. Volpe, and A. Camurri, "Analysis of intrapersonal synchronization in full-body movements displaying different expressive qualities," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ser. AVI '16. New York, NY, USA: ACM, 2016, pp. 136–143. [Online]. Available: http://doi.acm.org/10.1145/2909132.2909262

[32] A. M. Wing, S. Endo, A. Bradbury, and D. Vorberg, "Optimal feedback correction in string quartet synchronization," *Journal of The Royal Society Interface*, vol. 11, no. 93, p. 20131125, 2014.

[33] S.-f. Wong and K.-y. K. Wong, "Fast and reliable recognition of human motion from motion trajectories using wavelet analysis," in *Proc. the Symposium on Professional Practice in AI within the First IFIP Conference on Artificial Intelligence Applications and Innovations in IFIP World Computer Congress*, 2004.

[34] K. Fujiwara and I. Daibo, "Evaluating interpersonal synchrony: Wavelet transform toward an unstructured conversation," *Frontiers in psychology*, vol. 7, 2016.

[35] G. Varni, A. Camurri, P. Coletta, and G. Volpe, "Emotional entrainment in music performance," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–5.

[36] P. Alborno, G. Volpe, M. Mancini, R. Niewiadomski, S. Piana, and A. Camurri, "The multi event class synchronization algorithm," 2019. [Online]. Available: http://arxiv.org/abs/1903.09530

[37] T. Kreuz, D. Chicharro, R. G. Andrzejak, J. S. Haas, and H. D. Abarbanel, "Measuring multiple spike train synchrony," *Journal of neuroscience methods*, vol. 183, no. 2, pp. 287–299, 2009.

[38] A. Savitzky and M. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[39] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," *Discrete and Computational Geometry*, vol. 28, no. 4, pp. 511–533, 2002.

[40] G. M. Sullivan and R. Feinn, "Using effect size or why the p value is not enough," *Journal of Graduate Medical Education*, vol. 4, no. 3, pp. 279–282, 2012.

[41] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 1988.

[42] S. Sawilowsky, "New effect size rules of thumb," *Journal of Modern Applied Statistical Methods*, vol. 8, pp. 597–599, 11 2009.

[43] D. Kenny, *Statistics for the Social and Behavioral Sciences*. Little, Brown, 1987.

[44] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 3, no. 52, 1965.

[45] D. Lane. (2014) Online statistics education: An interactive multimedia course of study. [Online]. Available: http://onlinestatbook.com/2/calculators/normal_dist.html

Nikolas De Giorgis received a Ph.D. in Computer Science and Systems Engineering from the University of Genoa in 2018. His research interests include computer graphics, geometric and shape modeling, data visualization, geographic data processing and human-computer interaction. He is currently a Post-doctoral Researcher at the Italian Institute of Technology (Genova, Italy), working on 3D modeling of anthropometric measure from 3D laser scanners.

Enrico Puppo is professor of computer science at the DIBRIS, University of Genoa. He has been co-author of over 130 scientific publications on the subjects of spatial data handling, geometric modeling, geometry processing, computational geometry, and image processing. His current research interests are in geometry processing and multi-scale analysis, with applications to Computer Graphics, Computer Aided Design, Animation, Architecture, Scientific Visualization, and GIS.

Paolo Alborno received a Ph.D. in Computer Science and Systems Engineering from the University of Genoa. His research interests include intelligent and affective human-machine interaction, autonomous systems, modeling and real-time analysis and synthesis of expressive content, and multimodal interactive systems. He is currently a Post-doctoral Researcher at DIBRIS, University of Genoa part of the InfoMus - Casa Paganini Lab staff (www.infomus.org),

Antonio Camurri Ph.D. in Computer Engineering, is a full professor at DIBRIS, University of Genoa. Founder and scientific director of InfoMus Lab and of Casa Paganini (www.infomus.org). He is a coordinator and local project manager of several EU projects (FP5-FP7, H2020, Culture 2007, Cost Actions). His research interests include multimodal intelligent interfaces and interactive systems; sound and music computing; computational models of expressive gesture, emotion, and social signals; multimodal systems for music, dance, museums, health.