# D3.3 – EnTimeMent platform and software libraries for multi-time analysis, entrainment, and prediction - Phase 2

| | |
|---|---|
| Project No | GA824160 |
| Project Acronym | EnTimeMent |
| Project full title | ENtrainment & synchronization at multiple TIME scales in the MENTal foundations of expressive gesture |
| Instrument | FET Proactive |
| Type of action | RIA |
| Start Date of project | 1 January 2019 |
| Duration | 48 months |

| | |
|---|---|
| **Distribution level** | [PU][1] |
| **Due date of deliverable** | Month 24 |
| **Actual submission date** | January 2021 |
| **Deliverable number** | 3.3 |
| **Deliverable title** | EnTimeMent platform and software libraries for multi-time analysis, entrainment, and prediction - Phase 2 |
| **Type** | ORDP (Open Research Data Pilot) |
| **Status & version** | |
| **Number of pages** | |
| **WP contributing to the deliverable** | 3 |
| **WP / Task responsible** | UNIGE |
| **Other contributors** | ALL |
| **Author(s)** | UNIGE, Qualisys, UCL, UM-CN |
| **EC Project Officer** | Teresa De Martino |
| **Keywords** | Computational models, Software libraries, Movement analysis and prediction, Machine learning |

[1] **PU** = Public, **PP** = Restricted to other programme participants (including the Commission Services), **RE** = Restricted to a group specified by the consortium (including the Commission Services), **CO** = Confidential, only for members of the consortium (including the Commission Services).

Contents

1 Introduction 5

2 Activities in Phase 2 5

2.1 Datasets 5

2.1.1 Origin of Movement in Individual and dyad actions (UNIGE, EuroMov) – May–July 2020 5

2.1.2 Emotion dataset (Ellipsis) (UM-CN, UNIGE) – May-July 2020 ... 7

2.1.3 Markerless mocap campaign (UNIGE, Qualisys) - October-November 2021 ... 8

2.2 Hardware and Software platform modules 8

2.2.1 Overall architecture of the project platform at UNIGE ... 8

2.2.2 Sensor System for Chronic Pain Data Collection in Participant Homes ... 9

2.3 Software Libraries for individual movement features analysis 9

2.3.1 Software libraries for analysis of qualities of movement ... 9

2.3.2 Synchronization among temporal scales: the MECS algorithm ... 12

2.3.3 Automated measure of the origin of movement ... 13

2.3.4 Human Pose Estimation Architecture for computing Dyadic Synchronization ... 13

**Abbreviations**

| | |
|---|---|
| EU | European Union |
| EC | European Commission |
| WP | Work Package |

# 1 Introduction

This deliverable describes the updates (Phase 2) to the project technology platform developed in Year 2. The description of the project platform is available in D3.2, in this document we only describe the new hardware and software modules constituting the architecture of the project platform. The project technology platform is available online on the project repository.

# 2 Activities in Phase 2

## 2.1 Datasets

In this section we present novel datasets acquired in the second phase of EnTimeMent, using the project platform.

### 2.1.1 Origin of Movement in Individual and dyad actions (UNIGE, EuroMov) – May–July 2020

The experiment is in two parts

<u>PART 1 - Duo – Exchange of a ball</u>

One light ball (about 100g), and a heavier ball (about 2Kg). Same dimension of about 25cm diameter.

Two participants are located at a distance of 3m. They can move in a constrained rectangular space (an “island” 1x2m) identified by a visible tape on the floor to limit the island (1x2m).

<u>Action</u>: launch of the ball from one to the other using two hands. Using two hands should facilitate the involvement of the whole body in both launch and receive of the ball, and to avoid too high speed of launch resulting from one hand launch and to avoid complex movements like those of a baseball launcher.
Further, the launching and receiving action should not be symmetric: hands and feet positions not symmetric (e.g. one foot a little forward with respect to the other): asymmetry should contribute to “free” the movement from static postures.

Preparation (phase 0 – individual action):
Each actor separately launch the light ball to a “ghost” receiver in the other island.
Useful to compare how a launch alone differs from a launch in a duo.

Three modalities: The two participants are face-to-face, each standing in her own island.

2.1 “Fair” launch of the ball (cooperative, affiliative attitude): the two participants launch to each other the ball, trying to facilitate the grasp by the other. Both launch and receive are done using two hands.
Repeated launch-grasp at least 5 times for each participant (alternating).

2.2 “Hurt”: to throw the ball to maximise impact as if you want to hit/hurt the other person. The receiver gets no instruction. Launch and receive using two hands. Who launches performs with dominance; who receives has necessarily a defensive behaviour (fear-like).
Repeat at least 5 times for each participant (alternating).

2.3 “Cheating behaviour”: face-to-face, each moving in her own island, the sender launches the ball to the other trying to reduce the success of grasp by cheating actions, then the same is done by the other. Launch and receive using two hands. Repeat at least 5 times for each participant (alternating).

The order is always 2.1 - 2.2 – 2.3 (increasing difficulty).

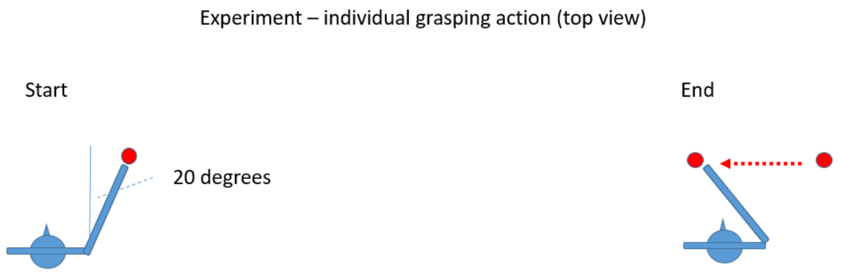
Research questions:
We look at the OoM of the sender and of the receiver.
This would raise nice issues about the time scales at which you perceive the OoM of your partner and change your own OoM.
Synchronization and anticipatory behaviour between sender-receiver.
Leader-follower behaviour in different emotions.
Does OoM contribute to explain different emotions.
What is the difference of doing an action (launch) alone or with another person.

*PART 2 – Individual action*

Objective: to detect the OoM in goal directed movements.
We start the movement from a similar rest posture (standing).
Objects: two glass bottles A and B, one empty and one filled with sand, covered by tape (identical at view).

Instructions:
The actor choose the bottle to start with: A or B (no information on the bottles)
Repeat for the two bottles starting from the one chosen:

— facing the bottle on a table
— distance shoulder-object = limit of reachable distance = object touched with outstretched arm and with closed fist (you can grasp the object without moving the shoulder when you open the fingers)
— direction of the object : 20° to the external direction from in front of my dominant shoulder
— displacement to : in front of my non-dominant shoulder
— height of the object : on a table top




We ask for a “spontaneous” reaching with the dominant hand, to grasp and move the object towards the non-dominant side

5 repetitions of the same experimental condition (to assess inter-trial variability).
We manipulate :

— the weight of the glass bottle (empty/full of sand): we know this should modify the body posture @grasp, and most likely the origin of movement.

Technical setup:
Mocap Qualisys, 16 cameras, Sport Markerset Qualisys
Two TV videocameras (front and side views) synchronized by SMPTE between them and with Mocap

Video excerpts from these multimodal recordings are available in the FETFX video:
http://www.fetfx.eu/news/entimement-new-video-timesregained/
The dataset is analyzed by project partners and a publication is planned. Then, the dataset will be made available on the project repository.

### 2.1.2 Emotion dataset (Ellipsis) (UM-CN, UNIGE) – May-July 2020

This dataset is based on a joint experiment of Maastricht University and UNIGE on the study of individual and dyad actions involving emotions. The multimodal recordings are organized in two phases. In the first session only video recordings are recorded. In the second session, video and mocap recordings are recorded. The video-only recordings will be used as stimuli in cognitive neuroscience experiments (avoiding the unusual clothing in mocap setups), the video and mocap recordings will be used for the analysis using computational models.

<u>First session – only video recording</u>
Clothing: no stripes and best is no or very little clothing difference between top and bottom halves. Take watches off, no colour shoes!

<u>Second session – video and mocap</u>
Each action repeated 5 times

Neutral actions:
1) Scratching: You are working in the kitchen, the doorbell is ringing, you quickly check yourself and you see that you have some flour on your pants, you brush it off. (5 movements, if right-handed they brush the left leg and vice versa).
2) Banana: you forget to have lunch, you are then very hungry, someone just gave you a banana (peel it 3 times and bite it once).
3) You are walking and you see a bush of blackberries: you bend over and pick 3 of them (one from your left, one from the centre and one from your right)
4) You need a specific tool located in a very low shelf, you squat down and just look for it (no gesture), you found it and stand up.
5) Somebody thrown a ball at you, you see it approaching and you get ready to catch it (don't approach the ball, just get ready to catch it).

Emotional actions:
1) You are walking, when you see a stone flying in your direction, you are forced to react suddenly to dodge it.
2) After the lockdown, you are free to walk in the street when you spot a good friend which you do not see since a lot, you start welcoming him/her since far away.

3) You feel your wallet is with you but then you realise you left it in a bar (react: you use your whole body and bang your arms)
4) Dominance: there is a bunch of teenagers/kids fooling around, you approach to make them stop (dominant posture)
5) Submission: your boss tells you that you made a stupid mistake and you accept that s/he s right.

Interactions (dyads, only same gender interactions):
1) Dominance/submission: A is the boss the B is the employee, recreate scenario of point 5 in emotional.
2) Grooming: A is talking to B, and B sees some dust on A's shoulder and brush it off
3) Verbal interaction (instead of fight): A and B simply talking with some gestures.
4) Greetings with the elbows: A and B meet and greet each other according to corona virus guidelines.
5) Feeding: A gives a banana to B and B starts to peel and eating it (3 peel + 1 bite).

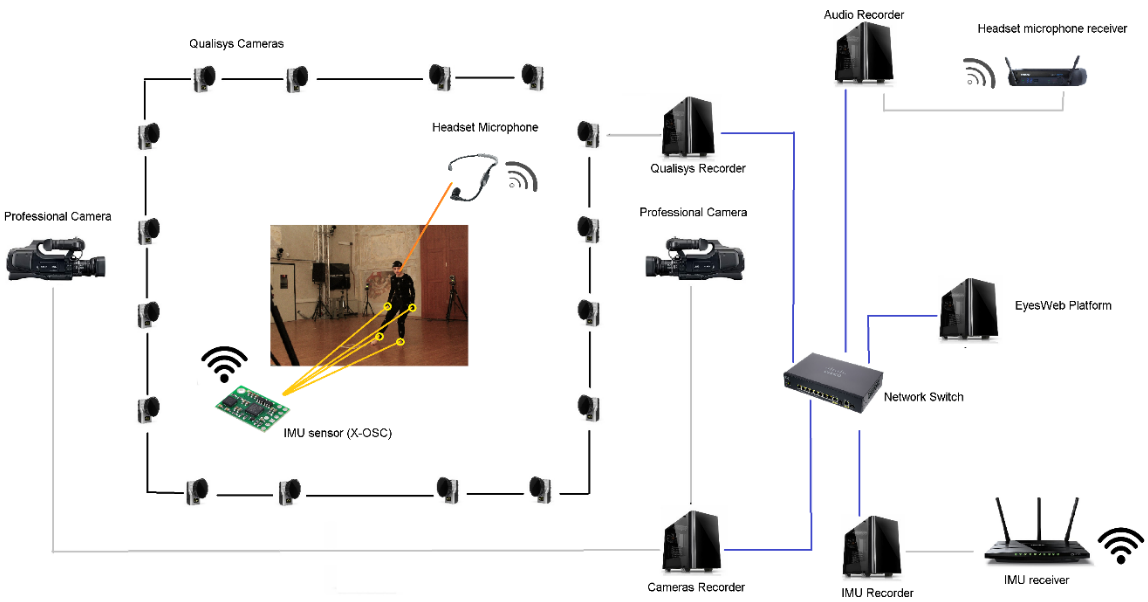
Technical setup:
Mocap Qualisys, 16 cameras, Sport Markerset Qualisys
Two TV videocameras (front and side views) synchronized by SMPTE between them and with Mocap

### 2.1.3 Markerless mocap campaign (UNIGE, Qualisys) - October-November 2021

A markerless motion capture campaign is planned as part of the Phase III evaluation of the scenarios in WP4 (see D4.10). The campaign will take place during a two-week period in autumn 2021. A markerless system will be made available for the occasion by Qualisys. The campaign gives the EnTimeMent partners the opportunity to explore the possibilities of markerless motion capture, an emerging technology that has promising potential for capturing of body movements in every-day settings and creative performance, where it may be difficult or intrusive to attach sensors to subjects. The exploration will allow the research partners to familiarize with the possibilities offered by this technology, and on the other hand provide useful feedback on the use of markerless motion capture in challenging settings to stimulate further development of markerless technology. The partners will submit ideas for experiments, allowing for an effective planning of the data acquisitions.

## 2.2 Hardware and Software platform modules

### 2.2.1 Overall architecture of the project platform at UNIGE




The recording platform is the same as in Phase 1 (see picture above), and it is composed by:

- 16 OQUS Camera (mixed setup with 700+, 700, 300)
- 2 Professional Cameras
- 1 Respiration Microphone for each participant
- 1-4 Accelerometers for each participant

The system allows to add biometric sensors or other devices. The architecture components are described in deliverable D3.2.
The platform standards adopted in Phase 1 remain confirmed, including the Qualisys Sport Markerset for motion capture.
A communication software library has been developed to connect EyesWeb and Qualisys Tracker Manager.
The library allows to receive streams from QTM both in real time or from a recorded file.
The EywEnTimeMent dll contains the following modules:

- QualisysSkeletonTSVReader: this module reads a TSV file exported from QTM, containing the skeleton sport joints coordinates and their rotations.
- QualisysTSVReader: it reads a TSV file exported from QTM, containing the markers coordinates and their rotations.
- QualisysRTProtocolDecoder: it decodes the Raw Data received from the QTM software via the OSC RealTime protocol. It can decodes both markers and skeleton

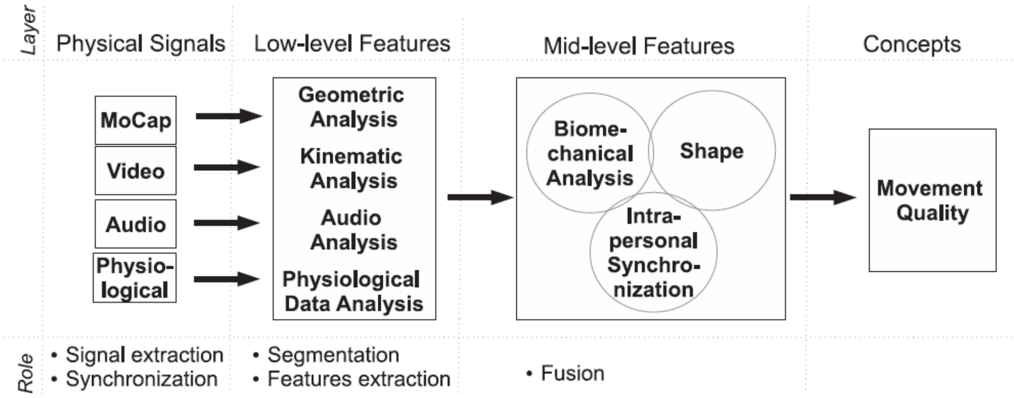
### 2.2.2 Sensor System for Chronic Pain Data Collection in Participant Homes

This platform component is described in D3.2.

## 2.3 Software Libraries for individual movement features analysis

### 2.3.1 Software libraries for analysis of qualities of movement

Our computational framework for the analysis of movement quality in full-body physical activities is described in D3.2 and is shown in the following figure




Our definition of movement quality focuses on the technical aspects of a movement performance. We exclude subjective factors that may influence the perception of movement quality at the individual level, for example, cultural background of the observer. Moreover, the correct performance of a movement (i.e., a high-quality movement) is usually related to the goals of the movement (e.g., an excellent performance in terms of intrabody synchronization may evoke positive aesthetic feelings in spectators).

In order to measure movement quality, we designed a computational framework consisting of four layers and several modules (see previous Figure). This is grounded on a conceptual framework conceived for analysis of expressive content conveyed by full-body movement and gesture (Camurri et al. 2004, 2016b).

The first layer of our framework–the *Physical Signals Layer*–consists of modules that capture and preprocess sources of data from different modalities. Below are more details about these modules.

—*MoCap Module*: It retrieves the 3D positions of body joints from a motion capture system, applies basic processing techniques (e.g., signal filtering), and computes basic kinematic features, such as velocity or acceleration.

—*Video Module*: It receives video streams from one or more video cameras and/or RGB-D sensors and possibly runs basic video-processing techniques (e.g., background subtraction and motion tracking).

—*AudioModule*: It captures audio streams from one ormore environmental or on-bodymicrophones and possibly runs basic audio-processing techniques (e.g., denoising). In our system, we focus on nonverbal audio.

—*Physiological Signals Module*: It retrieves data from physiological sensors — such as respiration or skin conductance response sensors — and applies basic processing techniques (e.g., signal filtering).

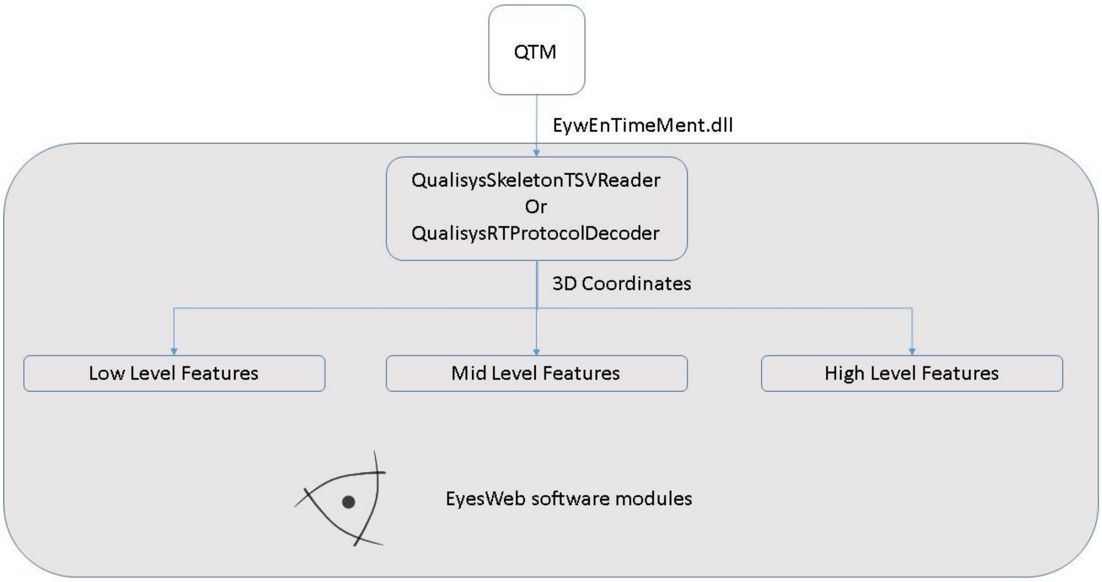
The *Low-Level Features Layer* consists of modules that compute basic features describing movement quality. Such modules are regrouped into different sets performing different analyses:

—*Geometric Analysis*, for example, computes distances and angles between joints.

—*Kinematic Analysis* computes movement trajectories as well as basic kinematic information (e.g., acceleration peaks and kinetic energy).

—*Audio Analysis* performs extraction and analysis of acoustic features, for example, Mel Frequency Cepstral Coefficients (MFCC) (Davis and Mermelstein 1980; Zheng et al. 2001), main frequency F0, or volume.

—*Physiological Data Analysis* performs physiological signal processing and analysis, for example, signal peak detection and signal periodicity.
At the *Mid-Level Features Layer*, quality is analyzed with respect to its three major components discussed in the previous section: *Biomechanical Analysis*, *Shape*, and *Intrapersonal Synchronization*.
Finally, on the top, the *Concepts Layer* is composed of one module that computes overall movement quality. The aim of this level is to merge the different facets of quality into one meaningful value. Since overall movement quality is related to the goals of each specific activity, different fusion models should be used for different activities. For example, Intrapersonal Synchronization may weight more in classic ballet, whereas Biomechanical Efficiency may weight more in sport activities that require a lot of physical effort.
This simple scheme represents the architecture and the behavior of the software for the analysys of movements.




A set of EyesWeb patches to analyze movements is under development and validation. The features are the following:

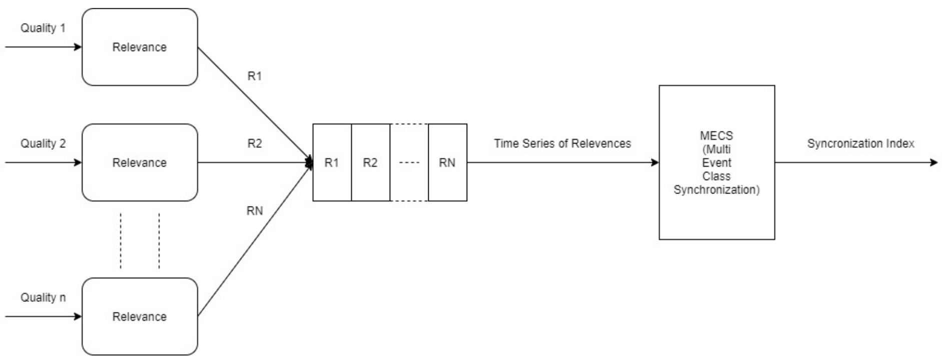
- Energy.
- Symmetry.
- Postural Tension
- Smoothness
- Fragility
- Lightness

For each quality, a specific temporal scale was selected, drawing from psychophysical studies on time perception and mental simulation (Fraisse 84).
After the extraction of the quality in a specific temporal scale, we evaluate the Relevance of the quality.
Relevance is an analysis primitive that can be computed on any movement quality *X*. The idea is to consider the histogram of *X* and to estimate the "distance" between the bin in which lies the current value of *X* and the bin corresponding to the most frequently occurring values of *X* in the "past".

Given the time series $x=x1,\dots,xn$ of $n$ observations of movement quality $X$ ($xn$ is the latest observation), Relevance is computed as follows:

- we compute HistX, the histogram of $X$, considering $\sqrt{n}$ equally spaced intervals; we call $occ_i$ the number of occurrences in interval $i$ ($i=1,\dots,\sqrt{n}$) of the elements of $x$,
- let iMAX be the interval corresponding to the highest bin (i.e., the bin of highest number of occurrences), and let occMAX be the number of occurrences in interval iMAX,
- let $i_n$ be the interval to which $x_n$ belongs to, and let $occ_n$ be the number of occurrences in $i_n$,
- we compute $D1=|iMAX- i_n|$,
- we compute $D2=occMAX-occ_n$,
- we compute Relevance as $D1{*}D2{*}$, where α is a constant positive real normalization factor.

A time series of the resulting different time scale relevances is analyzed by the MECS algorithm, to evaluate if exist a synchronization between the detected relevances




### 2.3.2 Synchronization among temporal scales: the MECS algorithm

Intra-personal synchronization is an important component in the analysis of the quality of individual movement. Inter-personal synchronization of movement qualities is important in the analysis of dyads and multi-person non-verbal social interaction.

*Multi-Event-Class Synchronization* (MECS) (Volpe, paper submitted) is a technique to measure synchronization between events detected in multiple time series. Synchronization is computed in terms of temporal alignment (within a time window) of the events occurring in the time series. After grouping events into classes, synchronization is computed within a class, i.e., between events belonging to the same class (*intra-class synchronization*) and between classes, i.e., between events belonging to different classes (*inter-class synchronization*). Additionally, events can be combined into *macro-events* on which synchronization is measured. A *macro-event* is an aggregation of events that satisfy some constraints. A relevant example of macro-event is a specific sequence of events. Events and macro-events can be grouped into *macro-classes* and synchronization can be computed within and between them.

In the framework of EnTimeMent, we expect to use MECS for measuring synchronization of events in time series of data characterizing movement at different time scales both at intra-personal and at inter-personal level. Moreover, the capability of MECS of handling macro-events can provide us with a useful tool to investigate synchronization at multiple time scales

since a macro-event at one time scale (e.g., a sequence of events at a low-level time scale) may correspond to one single event at another (higher-level) time scale.
The MECS algorithm has been improved with respect to the version described in D3.2 (Volpe, submitted).

*G. Volpe, "The Multi-Event-Class Synchronization (MECS) Algorithm", Submitted.*

### 2.3.3 Automated measure of the origin of movement

In (Kolykhalova et al., 2020) UNIGE proposed an approach to perform an automated analysis of the perceived origin of full-body human movement, i.e., the point at which such movement appears to be originated from the point of view of an observer. The approach is described in D3.2.
During the second phase of the project, this approach has been extended in collaboration with EuroMov, and presented at the ICMI 2020 EnTimeMent Intl Workshop (Matthiopoulou et al 2020). In particular, the nodes in the graph representing the body have been extended: besides speed, we considered acceleration and momentum. We are currently working at considering each node as a vector of features.
Possible further developments, include its application to a more complex skeletal structure (for which each cluster of joints is associated to a specific joint in the simpler 20-joint skeletal structure used in that work), making it possible to analyze movement in parallel at a finer interacting spatio-temporal scale in a multiple-scale approach (in line with the objectives of EnTimeMent). In this way, one could compare the Shapley value of a joint in the simpler structure with the sum of the Shapley values of the associated joints in the more complex structure (a smaller Shapley value would be expected for each of the latter joints).

*K. Kolykhalova, G. Gnecco, M. Sanguineti, G. Volpe, and A. Camurri (2020) "Automated analysis of the origin of movement: An approach based on cooperative games on graphs," IEEE Transactions on Human-Computer Studies, 2020.*

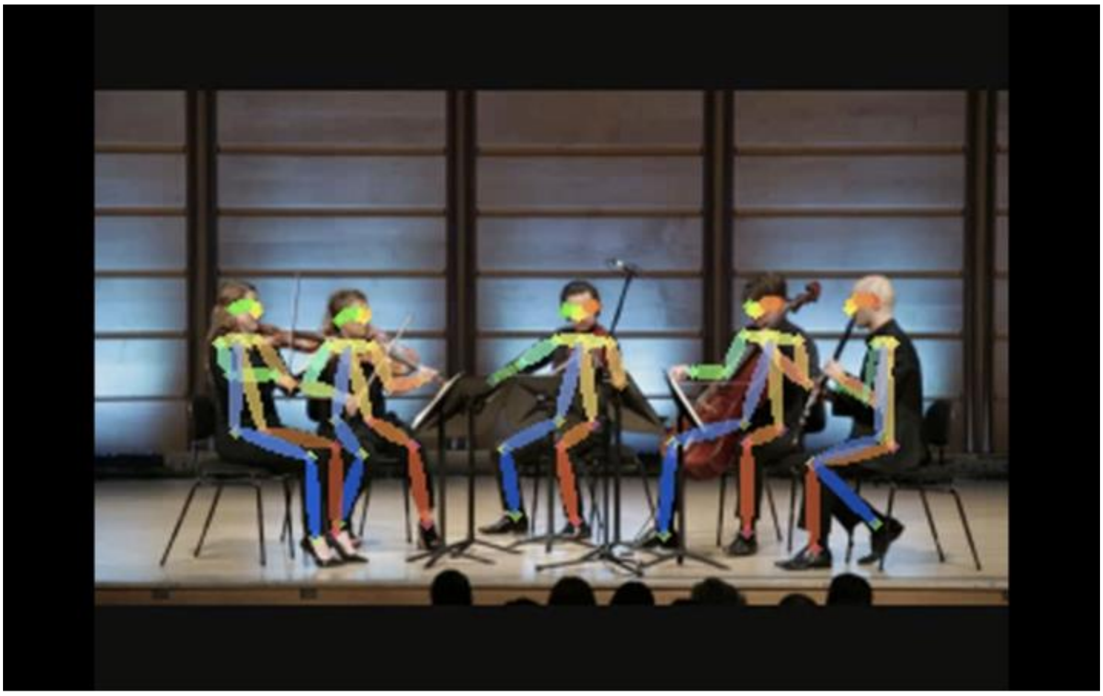
*O.Matthiopoulou et al. (2020) Automatic Detection of the Origin of Movement, ACM ICMI 2020 Entimement Intl Workshop.*

### 2.3.4 Human Pose Estimation Architecture for computing Dyadic Synchronization

In our study "Capturing human movement and shape information from small groups to extract expressive and social features – using marker-less techniques", we introduce a multi-person pose estimation technique to capture movement related data in a marker-less and non-intrusive manner.

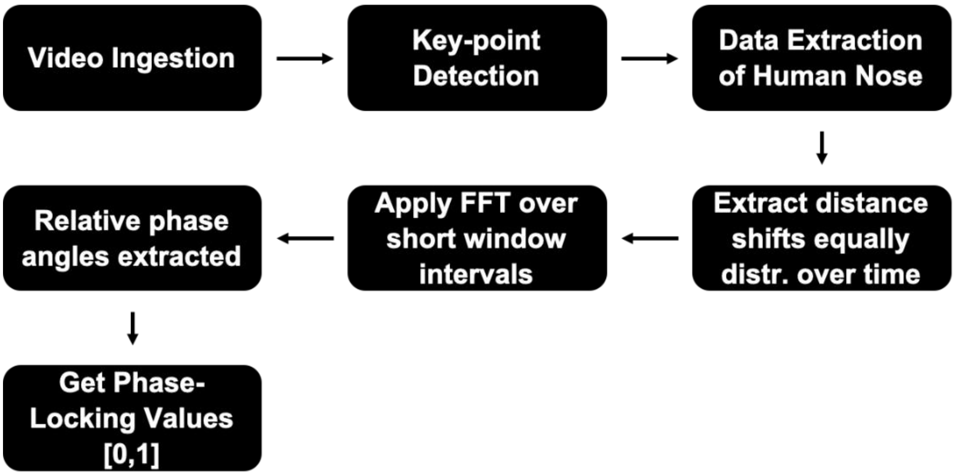
We make use of RMPE (Regional Multi-Person Pose Estimation) or AlphaPose (Fang, Hao-Shu, et al. 2017), which is a popular Multi-Person Pose Estimation algorithm. This algorithm utilises a Symmetric Spatial Transformer Network (SSTN) to extract a high-quality single person region from an inaccurate bounding box. A Single Person Pose Estimator (SPPE) is used in this extracted region to estimate the human pose skeleton for that person. A Spatial De-Transformer Network (SDTN) is used to remap the estimated human pose back to the original image coordinate system. Finally, a parametric pose Non-Maximum Suppression (NMS) technique is used to handle the issue of redundant pose deductions. A salient feature of AlphaPose is that this technique can be extended to any combination of a person detection algorithm, including a SPPE.

AlphaPose, or OpenPose (Cao, Zhe, et al. 2019) - another popular pose estimation algorithm developed at CMU - Carnegie Mellon University), are used to produce robust 2D Keypoint detections as an output in the form of a json file. This data can be extracted and exploited to study human movement behavior. These algorithms do pose their challenges as the noise produced is relatively high, but with time and advancements in training + prediction models, the overall accuracy is gradually increasing.



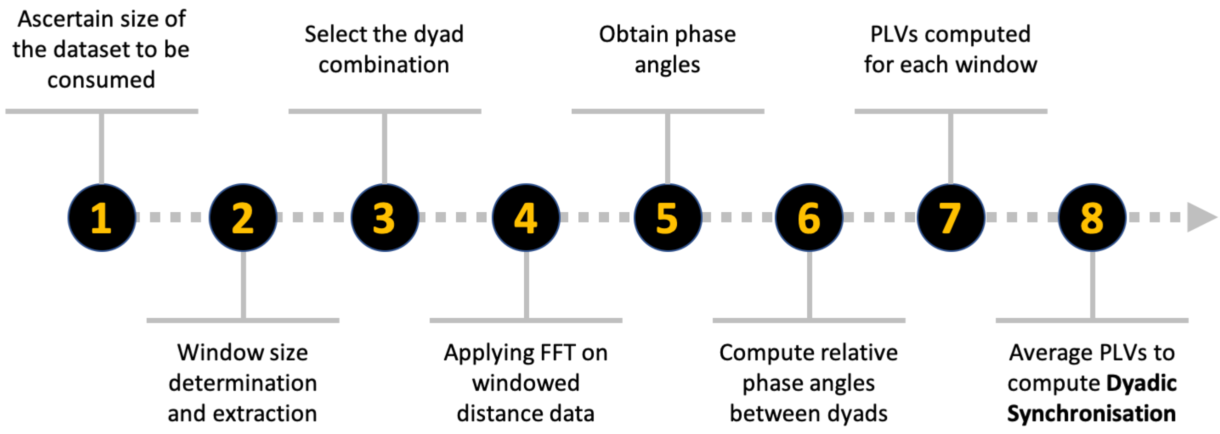
*An example of the output obtained using AlphaPose (a frame the Brahms Concert Part3) from the data repository of ensemble music performances of Western Sydney University.*

To investigate the feasibility of quantifying dyadic synchronization that exists between multiple performers in a Musical Ensemble, we pursue a 7-step methodology to achieve this.




*The Methodology followed for obtaining Phase-Locking Values*

By computing phase-locking values, we can compute Dyadic Synchronization that exists between two co-performers in a musical ensemble. Below is an illustration of the eight-step process pipeline that is utilzied to compute the Dyadic Synchronization.




*An illustration of the process pipeline for computing Dyadic Synchronization between a dyad, or in other words two co-performers.*

## References

*Fang, Hao-Shu, et al. "Rmpe: Regional multi-person pose estimation." Proceedings of the IEEE International Conference on Computer Vision. 2017.*

*Cao, Zhe, et al. "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields." IEEE transactions on pattern analysis and machine intelligence 43.1 (2019): 172-186.*