.

D1.1 Research requirements for laboratory and ecological experimental scenarios - Phase 1

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

TABLE OF CONTENTS

1. THEORETICAL FOUNDATIONS, COMPUTATIONAL MODELS AND ALGORITHMS

Understanding, measuring and predicting the qualities of movement imply a dynamic cognitive relation with a complex non-linearly stratified temporal dimension. Movements are hierarchically nested: a gesture sequence has a hierarchical layered structure - from high level layers down to more and more local components where every layer influences and is influenced by every other (bottom-up/top-down). Every layer is characterized by a different temporal dimension: a proper rhythm from macro to micro temporal scales of action. This organization does not only apply to action execution, but also to action observation and is at the basis of the unique human ability to understand and predict conspecific gestural qualities. Human skill in understanding and predicting gestural qualities, and attempting to influence one another's actions, depends on the capacity to create intercrossing relations between these different temporal and spatial layers through feedforward/feedback connections and bidirectional causalities, with the body as a time keeper, coordinating different internal, mental and physiological clocks. In 1973, Johansson showed that the human visual system can perceive the movement of a human body from a limited number of moving points. This landmark study grounded the scientific bases of current motion capture technologies. Recent studies proved that the information contained in such a limited number of moving points does not concern only the activity performed, but can also provide hints about more complex cognitive and affective phenomena: for example, Pollick (2001) showed that participants can infer emotional categories from point-light representations of everyday actions. Studies using naturalistic images and videos have established how fluent we are in body language (de Gelder, 2016). Very few studies consider the temporal dynamics of the stimulus, and how affective qualities may be perceived faster than other qualities (Meeren et al 2016), be interlinked and change over time. In other words, time is a crucial variable for these processes. Such time intervals are the time intervals of human perception and prediction, i.e., this is a human time, which integrates time at the neural level up to time at the level of narrative structures and content organization. Current technologies either do not deal with such a human time or they do in a quite empirical way: motion capture technologies are most often limited to computation of kinematic measures whose time frame is usually too short for an effective perception and prediction of complex phenomena. While a lot of effort is being spent improving such technologies in the direction of more accurate and more portable systems (e.g., wearable and wireless), such developments are incremental with respect to a conceptual and technological paradigm that remains unchanged. Furthermore, most systems for gesture recognition or for analysis of emotional content from movement data streams adopt time processing windows whose duration is fixed and is usually empirically determined.

Focusing on this last point, we can observe how these effects can be studied using several techniques. We want to create a mathematical model as accurate as possible, which is able to have predictions and able to understand actions performed by a complex system like the human one. To manage this type of problem from an analytical point of view, there are two different approaches:

1. Semi-empirical techniques are used to extract feature in order to define the state of the system and how it evolves over time. Since the features are derived from statistical measures (such as mean, variance and standard deviation), the operations that can be done using these data are very simple and limited to the scenario from which the data comes. Therefore it would be impractical to compare these features in different situations. Another consideration is that these techniques use time processing windows whose duration is fixed and is usually empirically determined. To use this type of approach is therefore too penalizing because a lot of information is lost, risking to apply only estimates on the behavior of a specific low level layer whereas, at higher level, a wrong prediction and understanding of

the action that is performed. Given the numerous limitations of this approach and the impracticality of their use in our project, it is clear that more advanced data analytic techniques are needed.

2. Advanced data analysis techniques allow a greater abstraction of the problem and therefore identify an optimal model useful for comparing results coming from different scenarios. Techniques often used are neural networks (NNs). These networks, once defined the basic architecture, and therefore the criterion with which the comparison is made, are able to be extended to similar problems. More complex features compared to simple statistical measurements are learned directly in the neural network training phase. However, NNs require a large amount of data to identify which features will be particularly useful for solving a specific task. Once these features are obtained these can be re-used to solve similar problems, thus allowing a greater abstraction of the problem. Therefore, it is clear how neural networks are a very powerful tool able to satisfy the themes of our project. As we have seen, it is sufficient to have a good number of data in order to have features that can be compared with each other by creating a model that is able to manage complex tasks such as the prediction of actions in hierarchical layered structures. Moreover, this type of structure allows an accurate analysis of the movements performed in an action, not simply analyzing estimates on the behavior of a given layer in its execution. Extensions of simple networks to more complex models such as Deep Neural Networks (DNNs) (Bengio, 2015), allows us to manage different time windows making it an extremely powerful tool.

An approach based on the use of statistical measurements is too limiting for the purpose of this project. In particular, as we have seen previously, semi-empirical techniques are too bound to the data from which features are extracted. Successively applying these features to action prediction tasks turns out to be an impractical choice for managing time sequences with hierarchical layered structure. Being this way based on simple estimates, a correct observation of the actions is particularly difficult. Moreover, these techniques use time processing windows whose duration is fixed and is usually empirically determined. In this way a loss of information is possible because some fragments of action can be omitted.

The best choice is therefore directed to the use of the second approach, where advanced data analysis techniques are used to have a more general abstraction of the problem addressed. In particular, we choose an approach aimed at (deep) neural networks so that features are learned in the training phase of the network itself and will be subsequently reused to handle similar situations. These features will be more complex than simple statistical measurements. As we have seen, however, a large number of data is needed to have increasingly more articulated and useful features for the task we want to solve. Furthermore, considering temporal sequences, it would be extraordinarily complicated (or even impossible) to apply approaches based on statistical measurements, analyzing only time window empirically determined. For the management of this type of problem, a subset of Deep Neural Networks capable of handling time sequences is used: recurrent neural networks. By analyzing the past information, these architectures allow an estimate of the future state. However, this information may not be sufficient: intuitively, to predict actions performed by a complex system, it will be necessary to predict all the components that make up the system itself in order to have a more accurate prediction. Therefore, it will be useful and significant to choose a model that includes the possibility of managing hierarchical layered structure of all the components that identify the system, each with its own temporal dimension.

A second consideration is due to the fact that usually the actions of a system are repeated or, simply, similar systems are able to perform similar actions. Then to predict an action of a complex system like the human one, the winning choice is linked to the use of a memory system that is able to understand events that are easily repeated over time. Therefore, we want to create a model that presents the advantages of recurrent networks and extends them by integrating a memory system.

1.1 Experiments, Scenarios and Objectives: from exploration to convergence

Our perspective is a human executing or observing a movement (e.g., a music or dance performance). We hypothesize a layered computational framework, from the physical low-level signals captured by sensors to the qualities – individual as well as social - that movement communicates, including emotions (Camurri et al 2016-MOCO intl conf). Movement at low (e.g., Motion Capture, EMG) layers is at a time scale of milliseconds, whilst wider time scales (e.g. NIRS, respiration) model higher layers. Time scales and layers are coexistent and mutually influence each other. An observer perceives salient expressive moments in a movement (e.g., a dance) both by its physical local low-level signals, and by its higher-level qualities, taking into account past events and emerging expectations: these, in their turn, change the observer's perspective and awareness of the low-level. That is, an observer of the movement changes her priorities and the importance within the large array of perceived physical signals. This interaction of processes at different time scales, as a continuous dialogue of coexisting parallel perspectives of the observed movement is a fundamental hypothesis of EnTimeMent, whose aim is to move towards a computational framework consisting of such different layers, ranging from physical signals to high-level individual as well as social qualities emerging from movement focusing on different space and time scales

The neuroscientific paradigm is based on the fundamental assumption that the cognitive experience of time requires a body. In fact, moving needs time and all our experiences, as well as their localization upon a reconstructed subjective experience of time, are dominated by the way we interact with our environment. The way we organize behaviour thus shapes the way we feel time and act according to its subjective representation. Human behaviour is indeed hierarchically organized in a way that each layer embraces a different time scale. Human behaviour is in fact constituted by goal-directed actions based on the synergic composition of simpler motor constituents chained together according to a precise and hierarchically organized "motor grammar" (Bernstein, 1967). In this view, the motor system can recombine or substitute motor elements to cope with a change in context, to achieve a new goal. Therefore, human natural experience coherently lives at the different scales characterizing human behaviour, at once.

The involvement of partners from different disciplines necessarily implies slightly different approaches and interests in the common questions of synchronization, entrainment, prediction, motor signature, empathy, and emotion at different temporal scales. In the first months of the project, significant effort has been on tackling these fundamental questions with a number of different feasibility studies and to the definition of protocols for both lab experiments (WP2) and applied scenarios (WP4). This work identifies the main directions of research foci for the following part of the project.

At this point of the project (M6), in Deliverable D1.1 all partners present tables describing Experiments or Research Programs fitting within the theoretical context of EnTimeMent. Some of these activities are in the early planning stages whereas others are in a more mature state. Importantly, as it will be evident in the following tables, many of the planned activities are shared among several partners. Convergence on a common framework will be pursued on data collection, analyses as well as the theoretical framework. In fact, the updated version of D1.1 due at M18, will include a more detailed description of the shared activities for each of the scenario and experimental type.

In conclusion, this Deliverable is to be considered as a living document describing the iterative process of convergence towards a small set of core theoretical questions, with all partners sharing the same conceptual framework and therefore experimental and analysis setups.

1.2 EnTimeMent data-sets

Availability of large high-quality data-sets is key to the definition of the goals defined by EnTimeMent. In this regard, a list of publicly available data-sets, including a brief description, has been collected and made available by UCL (Annex 1). At the same time, several partners in the consortium have already collected and are in the process of collecting new data. To facilitate collaborative research and to foster the development of shared research questions a list of data-sets will be provided in the form of tables.

2. PLANNED AND ONGOING RESEARCH ACTIVITY

2.1 Prediction in Action execution and observation

2.1.1 Bursty cortico-motor alpha coherence influence visual perception

Title	Bursty cortico-motor alpha coherence influence visual perception
Type	Experiment
Question of interest	The role and the non-stationarity properties of cortico-kinematic coherence in visual processing
Leaders	IIT-FE
Other ENTIMEMENT groups involved	None
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	The main objective is to study the role played by the rhythmic communication between the central nervous system and the periphery in driving visual perception beyond its role in motor performance.
Theoretical hypotheses	Cortico-motor communication works in irregular burst of intermittent communication which affects the active sampling of environmental information.
Operational hypotheses	We measure electroencephalographic data, movement kinematics in an isometric upper arm contraction. We intend to verify whether the emerging rhythmic communication between upper and lower motor centers affect perception.

Relationship with the objectives of the project	Upper and lower motor centers communicate at least according to two different time-scales below that of single movement - specifically at about 10 and 20 cycles per seconds. These constitute the basic time-scales affecting the sampling of sensory information during movement execution. This research will investigate these sensorimotor timescales.
Time schedule	Data collection terminated and analyses are ongoing.
Methods	TBA
Participants	25 healthy participants
Materials	Custom made isometric joystick. Electroencephalography (EEG).
Data format	Matlab data structure.
Experimental protocol/procedure	Each participant is required to maintain a stable isometric contraction for few second, while randomly in time, a threshold visual stimulus id presented to probe visual sensitivity.
Measures	Force transducers on the isometric joystick. Scalp electric potentials (EEG).
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.1.2 Motor recruitment during action observation: effect of interindividual differences in action strategy

Title	Motor recruitment during action observation: effect of interindividual differences in action strategy
Type	Experiment
Question of interest	Are individual motor signature (IMS) affecting action observation effects?
Leaders	IIT-FE
Other ENTIMEMENT groups involved	None
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	The coordination of our own actions with those of others requires the ability to read and anticipate what and how our partner is about to do. Indeed, when observing someone else moving, we can extract useful information such as future bodily displacements or infer higher-order

	cognitive processes hiding behind those actions. In principle, knowledge about the invariant properties of movement control could support inferences about the unfolding of other's actions.
Theoretical hypotheses	According to the predictive coding hypothesis, other's action sensory outcomes are compared to sensory predictions generated by the same hierarchical neural machinery for movement preparation and execution. This idea is however challenged by the redundancy that characterizes the organization of human movement. The abundance of degrees of freedom available during AE suggests that different joint configurations, as well as spatio-temporal patterns of muscle activity, can equally be used to reach the same behavioral goal. In this case, any sensorimotor-based inference about other's actions, amount to finding a solution to a many-to-many mapping problem.
Operational hypotheses	According to a strong version of the direct matching hypothesis, all subjects requested to observe the actions should mirror the muscle recruitment characterizing the actor. An alternative hypothesis predicts that motor activities would reflect, on an individual basis, a measure of the distance between own IMS and observed IMS. Furthermore, if sensorimotor activations are greater for little IMS distance, then it is likely that the motor system is computing the similarity between observed and own IMS. On the contrary, a negative relationship, would suggest that sensorimotor inferences about other's goals might be built by computing the difference or an error measure between one's own motor template and the observed movement.
Relationship with the objectives of the project	Perceptual discrimination and prediction of other's actions, may have a key role in supporting temporal and spatial interpersonal coordination. Here we suggest that a mapping exists between behavioral goals and the lower dimensionality space of whole-body configurations (i.e. synergies). On the top of that, everyone carry his own robust and yet unique way of moving (Individual Motor Signature – IMS). These two properties of human motor control may lead to a new one-to-one mapping that is function of everyone own way of moving (individual motor strategy, IMS). Backed by this, we hypothesize that while observing others' multi-joint actions, people build sensorimotor-based predictions by referencing what they see to the motor engrams of their own IMS.
Time schedule	Data collection finished. Data analyses running.
Methods	TBA
Participants	31 healthy participants
Materials	Electromyography, TMS, mocap.
Data format	Matlab data structure.
Experimental protocol/procedure	Participants first perform and then observe a whole-body reaching action which could be executed with different IMSs. After characterizing subjects' own IMS during execution, we measured their sensorimotor recruitment (corticospinal excitability, CSE) by administering single-pulse Transcranial Magnetic Stimulation (TMS) on their motor cortex while they observed an actor achieving the same goal by using different IMSs (i.e. the participant's own IMS and a different

	one). CSE was measured from the cortical representation of the Tibialis Anterior muscle (TA) that shows a clearly dissociable pattern while executing the two IMSs.
Measures	CSE; whole-body mocap.
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.1.3 Movement chain prediction in schizophrenic patients

Title	Movement chain prediction in schizophrenic patients
Type	Research Program
Question of interest	Are schizophrenic patient affected by problems in action anticipation?
Leaders	IIT-GE, IIT-FE
Other ENTIMEMENT groups involved	None
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	Schizophrenics patients have recently been described as having problems in timing-related tasks. Specifically, it has been proposed that some of their sub-clinical impairments resemble those of cerebellar patients that are characterized by fractioned action execution. Here we aim at understanding if these patients are also affected by a problem in other's action understanding.
Theoretical hypotheses	In this context, we aim at investigating one particular ability required for social interaction. Namely our ability to predict other's intentions. For example, any time a motor chain is activated (e.g., grasp-to-drink), the observer attributes the corresponding intention to the agent (e.g., drinking) from the first motor act (e.g., the grasp-to).
Operational hypotheses	In the current study, we investigate specific impairments, in the absence of discriminative contextual cues, in using slight kinematic variations in the observed grasp to inform mapping to the most probable chain.
Relationship with the objectives of the project	This study would describe a specific case of psychiatric impairment that extend its effect to a basic social skill, which is the ability to anticipate intentions of conspecifics.
Time schedule	Experiment in planning stage.
Methods	TBA

Participants	Schizophrenic patients (N to be defined) and a matched healthy control group.
Materials	Action video-clips, Behavioural responses.
Data format	Matlab data structure.
Experimental protocol/procedure	Participant will be submitted to an action observation experiment. From the dataset developed by Cavallo and colleagues, we will selected representative videos showing the reach to grasp phase of grasp-to-pour and grasp-to-drink actions. Each video clip will be presented at two levels of temporal occlusion (i.e. the video will stop at 25% or 100% of movement duration). Participant will have to discriminate the final intention.
Measures	Reaction times
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.1.4 Individual motor signature in weight-lifting task

Title	Individual motor signature in weight-lifting task
Type	Research Program
Question of interest	
Leaders	IIT-FE, UM-EuroMov
Other ENTIMEMENT groups involved	UNIGE
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	Describe the individual low-level specificity of movement control
Theoretical hypotheses	Each one of us move in the environment by planning ahead the coordination of a complex musculoskeletal system. Planning and execution of action must obey biomechanical and neural constraints and it is informed by past motor learning experience. All of this produce an individual motor signature.
Operational hypotheses	We intend to explore if in object lifting/moving there is an idiosyncratic weight-/mass kinematics relationship such that the gradual increase of weight/mass will be handled differently by each individual by scaling movement properties such as peak velocity or time to peak velocity.

	We plan to explore a moving object task (where the displacement is normal to the gravity field) and an object lifting task (where the displacement is parallel to the gravity field).
Relationship with the objectives of the project	This research activity has the scope of exploring the possibility to extract an individual motor signature from a simple and reliable task.
Time schedule	Ongoing
Methods	We record motion capture data while subject do an object lifting task. We manipulate spatial accuracy requirements and orientation with respect to gravity.
Participants	TBA
Materials	Movement position data, object acceleration and orientation
Data format	Matlab data structure.
Experimental protocol/procedure	Participants are requested to lift/move objects of the same size with different masses .
Measures	Movement position data
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

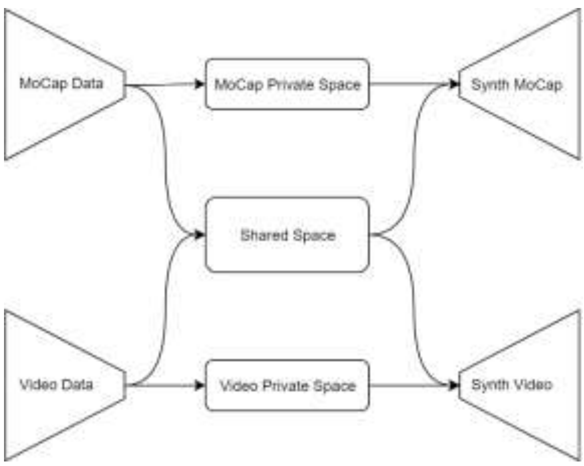
2.1.5 Motor equivalence in writing describe low-level individual motor signatures

Title	Motor equivalence in writing describe low-level individual motor signatures
Type	Research Program
Question of interest	Estimating presence and increase of different time scales for the same action performed with different intentions and/or effectors.
Leaders	UNIGE, IIT-FE
Other ENTIMEMENT groups involved	
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EnTimeMent in dancing with Times ● None of the above
Research objectives	1. Developing techniques for automated analysis of the presence of different time scales when the writing action is performed with different effectors.

	3. Correlating the results of the automated analysis with the results of perceptual ratings of the multiplicity of time scales.
Theoretical hypotheses	Executing the same action in different contexts and/or with different effectors, changes the relative relevance of different time scales contained in the action itself. Hence, both the intention of an action and its complexity reflect into the pattern of time scales.
Operational hypotheses	Actions performed with different effectors maintain the same proportionality across time-scales.
Relationship with the objectives of the project	Exploring the spatial scale-invariance of actions by analysing the data at multiple time-scales at the same time.
Time schedule	Early pilot data collection and ongoing planning of experiments
Methods	Certain kinds of recurrent neural networks, such as the Clockwork Recurrent Neural Network (CW-RNN), have demonstrated to be able to work well with time series associated with different time scales. Still, assessing the importance of recurrent neural network modules associated with different time scales is an open problem. In a second phase of this research program, the pattern of relevant time scales might be estimated by combining Recurrent Neural Networks (RNNs) and Cooperative Game Theory. As regards the former, the Clockwork RNN(CW-RNN) and its variations will be considered. The network is made of g modules of hidden neurons. Each module i is associated with a different period T_i, whose purpose is to capture a different time scale. “Faster” neurons (associated with smaller T_i’s) receive inputs from “slower” neurons (associated with larger T_i’s), and their weights are updated through back-propagation more frequently. Different modules may have different importance for different tasks (e.g., for certain “simple” tasks, the “slowest” neurons may be enough to get a satisfying performance). The CW-RNN will be trained via a data set obtained from the chosen action. Then, a Cooperative Game with Transferable Utility, called Clockwork Recurrent Neural Network Game (CW-RNN-G) will be defined on the trained network, such that: (i) the players are the network modules; (ii) each coalition of players corresponds to a different architecture of the CW-RNN, containing only the respective modules; (iii) the utility of coalitions is defined and computed in the following way: a. for each coalition, the network is trained using the training set; b. the coalition utility is the accuracy of the trained network computed on a validation set. Since the goal here is to assess the importance of different modules, it would be fair to re-train the network for each coalition. However, to save computational time, one may try to avoid a complete re-training. A pre-training phase could be also performed.

	The game-theoretical concept of “Shapley value” will be used in the CW-RNN-G to estimate the relative importance of different time scales. The Shapley value of each module represents its average marginal contribution to accuracy, when it is inserted in a random coalition of modules. (iv) The vector of computed Shapley values could be used to define a measure of similarity of the execution of the action with different intentions or effectors. (v) Such measure of similarity could be the Kendall’s tau correlation coefficient of the modules rankings obtained for different tasks. (vi) As an alternative, the measure of similarity could take into account the number of modules whose relative Shapley value is above a suitable threshold. (vii) The outcomes of this similarity analysis could be exploited to recognize and cluster actions performed with similar intensions or effectors. (viii) A subjective evaluation of suitable features associated with the task (e.g., their in terms of number of time scales involved, and the importance of different time scales for the specific task) could be used to validate such measure of similarity. This could be done via a suitably-designed online survey. At the end of the analysis, statistical tests could be applied to assess the statistical significance of the results. From a computational point of view, Monte Carlo sampling could be used to get approximations of the Shapley values, when a large number of modules is present.
Participants	TBD
Materials	Material: -Synchronized Audio/Video/MoCap recordings
Data format	Matlab data structure.
Experimental protocol/procedure	Participants will produce writing action in several condition. On paper on a board in the air, with the whole arm and with the head. The scope is to extract an individual spatial-scale independent kinematic fingerprint.
Measures	Automated multiple time scales analysis. Participants’ ratings
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.1.6 Multi-Cue Movement Analysis using a Shared Representation

Title	Multi-Cue Movement Analysis using a Shared Representation
Type	Research Program
Question of interest	Can a shared latent representation be learned between multiple cues, so that data can be transferred between cues to fill in gaps in observations?
Leaders	KTH
Other ENTIMEMENT groups involved	None
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	The main objective is to analyse human movements based on multiple cues, such as MoCap, video and IMU data, but allow some data to be missing. When data is missing, for example, when we only have video data, a shared representation will be utilized to synthesize MoCap data, using a framework similar to the bimodal deep autoencoder shown below. 
Theoretical hypotheses	Some aspects of human movements are shared by multiple cues and can be captured by a shared representation, whereas others are cue specific and need a private representation for each cue to be fully reconstructed.
Operational hypotheses	A deep autoencoder structure, such as the one shown above, that includes a shared latent representation and private cue specific representations allows transfer of data from one cue to another.
Relationship with the objectives of the project	This experiment relates to Task 3.4: short-term gesture prediction. It will test the possibility of finding a shared latent representation from

	multiple cues and use this representation for prediction in movement qualities over different time scales.
Time schedule	Data collection, method development and analysis will be completed in the ENTIMEMENT project.
Methods	TBA
Participants	TBA
Materials	We will collect data under the scenarios such as one-on-one basketball and human subjects engaged in domestic work.
Data format	MoCap skeleton data, 3D skeleton / full-body positions obtained from video, video data, possible RGB-D data; The human activities should be specific enough, including all kinds of movements, such as arm wave, high arm wave, hand catch, throw, hand clap, kick, walking, etc.
Experimental protocol/procedure	TBA
Measures	Motion Capture: 1. Use Qualisys MoCap to capture full body skeletons. 2. Use Kinect V2 or other videocameras to capture data in other formats.
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.1.7 Movement qualities in musical performance

Title	Movement qualities in music performance
Type	Research Program
Question of interest	Exploring interactions between movement qualities at different time scales in musical behaviour, with reference to expression, interpersonal interaction and performance regulation
Leaders	DU
Other ENTIMEMENT groups involved	UNIGE, UM-EuroMov

Experiment type (see WP2)	· Task2.1: Prediction in Action execution and observation · Task2.2: Prediction in Dyadic Action execution and observation · Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EnTimeMent in dancing with Times · None of the above
Research objectives	In line with the aims of Task 1.3, we plan to build on the insights and data collection of the IEMP project by exploring the movement qualities of musical performers at different time scales. The objective is to explore the IEMP corpus of North Indian Raga performances to understand the relationship of individual performers' movements to musical/ gestural phrases typical of Indian modes (raga), to prescribed metrical structures (tala), and to the management of performance (including interactions between the movements of different performers at both synchronisation (100-2000ms) and coordination (>10s) timescale).
Theoretical hypotheses	We hypothesize that it will be possible to recognise the salience of an individual's movements by establishing the typical movement qualities associated with (i) beat markers, (ii) cadence markers, (iii) melody accompaniment (e.g. tracing, pointing), and (iv) intention to interact with others. Other factors such as changes in timbre or dynamics may also be relevant. Movements associated with expression (for example of a specific mood or emotion) should be associated most strongly with the third category (melody accompaniment) and related to similar movement qualities in 'real life' emotional expression.

Operational hypotheses	Manual annotation of musical contents, gesture content and reference, structure, and interpersonal interaction will be combined with audio information and upper-body movement data extracted using the OpenPose system. Analysis will explore which audio and movement features, at which time-scale, predict which annotated factors. The aim is to establish predictors for movement salience (i.e. when movement indicates a beat, when it indicates expressive content), the identity of individual musicians or the identity of the musical mode (raga). We have shown previously using cross-wavelet transform analysis how coherence between musicians' movements can be a predictor of structural transitions in the music, presumably because they pay more mutual attention at these moments. A further possible extension of the programme will be to ask participants to respond to musical excerpts with instructions such as 'try to trace the melody', and then ask for feedback on the examples (e.g. emotional content). This would allow us to explore the extent to which expressive movement qualities are effectively encoded in the audio.

Relationship with the objectives of the project	The objective of further developing insights from IEMP to explore interactions between different time scales in music performance, is set our in Task 1.3. Sonification of movement forms a part of several strands of EnTimeMent. Insights from this work on detecting **salience** of different kinds of performer movement can be applied in that work. By using extended recordings of complex actions and interactions in small groups (2-5 people), this work provides insights into the development of interpersonal **interaction** and the mutual influence of movement patterns at different time scales. Exploring the 'motor signatures' of specific musical repertory items and their typical melodic movements will allow this information to be integrated with qualitative annotations and interview data about those items concerning the imagined movements, characters, moods and emotions with which they are associated. This therefore allows exploration of the way in which music, movement and **expression** are interrelated. Indian singers often comment that their gestures should look 'natural', and it is often remarked that they can relate to physical actions such as drawing a thread, stretching an elastic band or transferring a weight. The collaborations in this project allow us to explore the relationship of such virtual object-manipulation to real actions and object manipulations. It also allows us to explore specific movement qualities in terms of **responses to gravity**. For instance, do gestures indicate that ascending melodies must work against gravity, descending melodies with its help? How do beats utilise gravity?

Time schedule	Extraction of movement data and extension of annotations of performances from the IEMP NIR collections: from autumn 2019. Analysis from early 2020. New recordings from early 2020 according to needs determined in the preliminary analysis: e.g MoCap recordings for comparison with OpenPose data; recordings with additional physiological element; recordings of participants responding to music extracts.
Methods	Extraction of musicians' movement from video using OpenPose system. Manual annotation of recordings to complement and enrich existing annotations. New methods to be developed using machine learning techniques to explore the prediction of annotation categories from multimodal data input. New multimodal data recordings will be made with Indian musicians performing extracts from specific ragas (modes) and moving in response to music excerpts. Recordings will use audio, video, motion capture, physiological markers (e.g. ECG, respiration).
Participants	Indian musicians (and possibly dancers)

Materials	Existing materials from the IEMP and linked projects (Durham holds a much larger collection from which to draw more examples). By autumn 2019 will include 17 raga performances, 12 vocal + 5 instrumental. We will add more recordings to the annotated collection according to need. New materials. Musicians may be asked to perform short solo pieces in a number of specified North Indian ragas or talas. These will be long enough to include the main features (e.g. melodic movements, ornaments, typical drum patterns) and include moments of initiation, emphasis and cadence. Listeners will be asked to respond to audio recordings of extracts from the same recordings. For beat marking studies, materials would include metronome clicks/beeps, generic stylistic drum loops at different tempi, and examples of real music: all of these are easily available.
Data format	WAV audio, MP4 video Movement data and annotations CSV
Experimental protocol/procedure	Performance examples: expert musicians will be asked to perform short pieces related to those analysed from performances, to allow us to explore the interaction between individual movement style and repertoire-specifc movement; and between solo and accompanied movement. Response experiments: individual listeners, some of whom will be trained musicians or dancers, will be played audio excerpts from the analysed recordings and asked to move with the music. Instructions may be either beat-specific ("Try to indicate the beat of the music") or melody-specific ("Try to trace the melody with your hands").
Measures	Motion capture (musicians' hands, heads and shoulders), video, audio, EMG, respiration, observer perceptual judgements and expert annotations.
Results	

Descriptive results	
Inference statistics	
Additional results	
Discussion	

2.1.8 Generative Models for Movement Generation to Facilitate Social Interaction

Title	Generative Models for Movement Generation to Facilitate Social Interaction
Type	Research Program
Question of interest	Can an avatar (e.g. a projected silhouette of a moving person) driven by a generative model learned from observing human examples, express emotional states though movements to facilitate interaction with a human partner?
Leaders	KTH
Other ENTIMEMENT groups involved	
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ● None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ● Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	The main objective is to synthesize movements through a multi-stage process based on generative models, to make an avatar react to the movements of a human partner and express emotional states.
Theoretical hypotheses	Movements can be generated by generative models to express different emotions or other qualities. Such expression can be used to increase the degree of social interaction.
Operational hypotheses	A latent space representation of human movement can be learned, where some dimensions are forced to capture emotional states. This representation can then be used by a generative model to create a silhouette of a moving person for which the emotional state can be controlled. The movement of the silhouette can be adapted from observing the response of the human partner.

Relationship with the objectives of the project	This study relates to Task 3.6: Motion generation for social interaction.
Time schedule	Data collection, method development and analysis will be completed in the ENTIMEMENT project.
Methods	TBA
Participants	TBA
Materials	Videos of human dancers and actors expressing different emotional states, with corresponding annotated silhouettes.
Data format	RGB video data , binary images of silhouettes
Experimental protocol/procedure	TBA
Measures	TBA
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.1.9 Multi-time ML techniques for movement prediction

Title	Multi-time ML techniques for movement prediction.
Type	Research Program
Question of interest	To investigate ML techniques to determine the dimensionality of temporal scales to predict human movement in individual scenarios .
Leaders	KTH, UNIGE
Other ENTIMEMENT groups involved	
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ● None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ● Task 4.3: Scenario 3 - EnTimeMent in dancing with Times ● None of the above
Research objectives	Evaluating Neural Network models to explain EnTimeMent phenomena at different time-scales.
Theoretical hypotheses	Improvised movements performed with different qualities like lightness or fragility might need different time-scales.
Operational hypotheses	We start from multi-timescales machine learning methods, including CW-RNN, MT-LSTM, Autoencoder.
Relationship with the objectives of the project	This study relates to Task 3.5.
Time schedule	Start July 2019 to study models and to choose the dataset.

Methods	CW-RNN, MT-LSTM, Autoencoder
Participants	12 dancer
Materials	
Data format	VIDEO; IMU
Experimental protocol/procedure	TBA
Measures	TBA Lightness and Fragility IMU and video dataset
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.1.10 Understanding Movement Assessment Timescales

Title	Understanding Movement Assessment Timescales
Type	Research Program
Question of interest	What temporal segments do movement experts (e.g. physios) base their assessment of movement data on (e.g., of patients)?, Are different temporal scales helpful for different aspects of movement and related states?, Can we use this understanding to improve machine learning performance?
Leaders	UCL
Other EnTimeMent groups involved	None
Experiment Type (see WP2)	☒ Task 2.1: Prediction in Action execution and observation ☐ Task 2.2: Prediction in Dyadic Action execution and observation ☐ Task 2.3: Prediction in Complex Action execution and observation ☐ None of the above
Use Case Scenario (see WP4)	☐ Task 4.1: Scenario 1 - Healing with multiple times ☒ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ☐ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ☐ None of the above
Research objectives	● To understand how the machine-learning-based attention distribution varies with multiple timescales ● To understand the temporal scales in which clinicians assess body movement of people with chronic pain
Theoretical hypotheses	None
Operational hypotheses	None
Relationship with the objectives of the project	Aims to contribute: ● a machine learning architecture for modeling movement at multiple timescales ● to the understanding of human perception of movement qualities in relation to pain

	●
Time schedule	from June 2019
Methods	Machine Learning; Possibly Video Analysis
Participants	Possibly Physiotherapists
Materials	EmoPain motion capture data
Data format	Motion capture sequences
Experimental protocol/procedure	● build attention-based machine learning algorithms (e.g. BANet) and adapt BANet to timescale-BANet to allow timescale manipulation ● analyse model attention scores ● possibly also get physiotherapist analysis of videos for more in depth exploration
Measures	TBA
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.1.11 Exploring Multiscale Event Segmentation

Title	Exploring Multiscale Event Segmentation
Type	Research program
Question of interest	How can movement data be auto-segmented at multiple timescales?, What temporal segments of movement (from multiple timescales) map to relevant cognitive/affective experiences?, How can these segments be auto-mapped to these labels? How can we create motivic ('memorable') music (small scale) from movement, that sit on the sonification/music segments boundaries?, How can we integrate these motives into larger-scale forms?, Based on the developed computational segmentation models, can we create musical trajectories that reflect action trajectories, such that the motives occur at movement segment boundaries?, Can this musical framework be used to provide recall cues at a later time?
Leaders	UCL
Other EnTimeMent groups involved	None
Experiment Type (see WP2)	☐ Task 2.1: Prediction in Action execution and observation ☐ Task 2.2: Prediction in Dyadic Action execution and observation ☐ Task 2.3: Prediction in Complex Action execution and observation ☒ None of the above
Use Case Scenario (see WP4)	☐ Task 4.1: Scenario 1 - Healing with multiple times ☒ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ☐ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ☐ None of the above

Research objectives	● to explore the possibility of auto-segmenting movement data into events at multiple time scales, where event boundaries mark changes in movement (e.g. a new action, or activity) and/or changes in higher level semantics of movement (e.g. cognitive or affective experiences) ● to understand the feasibility of creating motivic music (i.e. small scale), from movement, that sit on the sonification/music segments boundaries, explore whether these can be developed into larger-scale forms to create musical trajectory that reflects action trajectory based on the computational segmentation models from machine learning studies, such that the motives occur at movement segment boundaries
Theoretical hypotheses	NA
Operational hypotheses	NA
Relationship with the objectives of the project	Aims to contribute: ● a machine learning architecture for modeling movement at multiple timescales ● a multi-timescale sonification (framework)
Time schedule	from January 2020
Methods	Data Collection; Machine Learning; Sonification
Participants	Healthy People; People with Chronic Pain
Materials	EmoPain motion capture data
Data format	Motion capture data
Experimental protocol/procedure	● build machine learning architecture ● collect data ● develop novel sonification approaches to possibly alter movement perception and execution
Measures	TBA
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.1.12 Prediction of visual perception related brain activity by kinematic and postural movement features

Title	Prediction of visual perception related brain activity by kinematic and postural movement features
Type	Research Program
Question of interest	What features of body movement drive activity in body perception related brain regions?
Leaders	UM

Other ENTIMEMENT groups involved	UNIGE, ITT-FE
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	To establish a set of body movement features that can explain and predict brain signals from regions in the visual processing hierarchy responsible for body perception and movement decoding.
Theoretical hypotheses	Humans are able to understand, interpret and predict visual input from human motion with apparent ease and high accuracy. It is not clear still how the human brain solves this task. The hypothesis is that the brain decomposes the visual input at different levels into internal representations that encompass spatial and temporal scales going from fine to coarse and that these representations are maintained in distinct brain regions.
Operational hypotheses	There is not a single brain region responsible for body perception, rather a set of hierarchical organized areas cooperate to form an understanding of the perceived body and it's motion. We hypothesize that there is a correspondence between the activity of single regions and a level of description in terms of computational movement features, such that the activity of said regions in response to a visual stimulus can be predicted based on a combination of features derived from the stimulus.
Relationship with the objectives of the project	This study provides information on how the human brain tackles the task of understanding body movement at different time scales.
Time schedule	Experiment in planning stage.

Methods	fMRI, computer vision, image and statistical analyses
Participants	Healthy participants
Materials	Human body motion video-clips, behavioural responses, fMRI data
Data format	Matlab and python data structures.
Experimental protocol/procedure	Participant will be scanned in an MRI while watching the stimuli developed for this research program.
Measures	Brain activity as measured by fMRI
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.2 Prediction in Dyadic Action execution and observation

○

2.2.1 Dyadic coordination of sub-movements

Title	Dyadic coordination of sub-movements
Type	Research Program
Question of interest	Are sub-movements contagious as we know movements are?
Leaders	IIT-FE
Other ENTIMEMENT groups involved	None
Experiment type (see WP2)	○ Task2.1: Prediction in Action execution and observation ● Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times

	○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	The main objective is to study if dyadic coordination affect sub-movements expression and coordination
Theoretical hypotheses	Sub-movements in the range of 2-4 Hz have been described to be affected by visual feedback during action execution. We intend to verify if action coordination contaminate the expression of these discontinuities present in (slow) visually-guided actions.
Operational hypotheses	We measure movement kinematics in a finger flexion-extension action in a solo and dyadic condition (in phase and anti-phase). We intend to verify whether the sub-movement rhythmicity is affected by the interaction.
Relationship with the objectives of the project	Sub-movements have recently been proposed to be mostly generated by passive peripheral resonance mechanisms. If we show that behavioural coordination produces automatic kinematic contagion across partners, we will first demonstrate a cortical origin for sub-movements while at the same time we would extend the phenomena of automatic imitation to a finer timescale of action execution.
Time schedule	Data collection ongoing.
Methods	TBA
Participants	40 healthy participants
Materials	Mocap
Data format	Matlab data structure.
Experimental protocol/procedure	Each participant is required to produce rhythmic index finger flexion-extension movements, alone or in coordination with a partner.
Measures	Movement kinematics
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.2.2 Motor activations during concurrent action execution and observation

Title	Motor cortical inhibition during concurrent action execution (AE) and action observation (AO)
Type	Research Program
Question of interest	Are AO effects modulated by concurrent AO?
Leaders	IIT-FE
Other ENTIMEMENT groups involved	None
Experiment type (see WP2)	○ Task2.1: Prediction in Action execution and observation ● Task2.2: Prediction in Dyadic Action execution and observation

	○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	Action Execution (AE) and Action Observation (AO) share an extended cortical network of activated areas. During coordinative action these processes also overlap in time, potentially giving rise to behavioral interference effects. The neurophysiological mechanisms subtending the interaction between concurrent AE and AO are substantially unknown.
Theoretical hypotheses	According to the predictive coding hypothesis, other's action sensory outcomes are compared to sensory predictions generated by the same hierarchical neural machinery for movement preparation and execution.
Operational hypotheses	We designed four experiments, to elucidate the neurophysiological mechanisms subtending the integration of AO and AE. Participants were asked perform an action, while observing the same or a different action. The dependent measure was the length of the Cortical Silent Period (CSP) elicited from the FDS muscle. CSP is a GABAb-mediated corticospinal index of inhibition associated with the voluntary motor drive and regarded as a marker of response selection.
Relationship with the objectives of the project	Perceptual discrimination and prediction of other's actions, may have a key role in supporting temporal and spatial interpersonal coordination. We may indeed observe other's actions, to produce complementary responses in a turn-taking fashion (e.g., playing tennis) or to simultaneously coordinate our own movements with those of others (e.g., when moving a heavy object together). However, the cortical response to new stimuli is influenced by ongoing activity in the same neural substrate. We can thus expect that temporal and spatial overlap of the neural processes subtending AE and AO produces functionally relevant interaction.
Time schedule	Data collection finished. Data analyses running.
Methods	TBA
Participants	64 healthy participants
Materials	Electromyography and TMS.
Data format	Matlab data structure.
Experimental protocol/procedure	In the main transcranial magnetic stimulation (TMS) study, participants were asked to keep the same isometric opened or closed hand posture, while observing an intransitive hand opening or closing action.
Measures	CSPs
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.2.3 Anticipatory postural adjustments (APA) during joint action coordination

Title	Anticipatory postural adjustments (APA) during joint action coordination
Type	Experiment
Question of interest	Are APAs triggered during dyadic action?
Leaders	IIT-FE
Other ENTIMEMENT groups involved	None
Experiment type (see WP2)	○ Task2.1: Prediction in Action execution and observation ● Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	It is recurrently claimed that human effortlessly detect others' hidden mental state by simply observing their movements and transforming the visual input into motor knowledge to predict their behavior. Using a classical paradigm quantifying motor predictions we tested the role of vision feedback during a reach and load-lifting task performed either alone, or with the help of a partner.
Theoretical hypotheses	We intend to show whether during dyadic interaction, in addition to self-motor representations, individuals adapt the cooperation by continuously integrating sensory signals coming from various sources.
Operational hypotheses	Wrist flexor and extensors muscle activities were recorded on the supporting hand. Early muscle changes preventing limb instabilities when participants performed the task by themselves, revealed the contribution of the visual input in postural anticipation. When the partner performed the unloading, a condition mimicking a split-brain situation, motor prediction followed a pattern evolving along the task course and gaining from the integration of the successive somatosensory feedbacks.
Relationship with the objectives of the project	Perceptual discrimination and prediction of other's actions, may have a key role in supporting temporal and spatial interpersonal coordination. Here we intend to verify whether visual action prediction affect low level control parameters such as the one instantiated by APAs and thus related to maintaining postural equilibrium.
Time schedule	Data collection finished. Data analyses running.
Methods	TBA
Participants	34 healthy participants
Materials	Electromyography

Data format	Matlab data structure.
Experimental protocol/procedure	The two participants sat face-to-face separated. In each couple, one participant was designated as the "Carrier", and the other as the "Partner". In a first experimental condition, the carrier performed the task by her/himself (Self condition) by holding the tray with his left hand while reaching, grasping and lifting the object with her/his right hand. In a second experimental condition, the partner had to reach, grasp and lift the carrier's object with his right hand (Joint condition). These two conditions were carried out with the carrier having either the eyes open (EO) or closed (EC).
Measures	Carrier's arm flexor/extensor EMG onset with respect to object touch and lift.
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.2.4 Representing Human Movement in Dyadic Actions over Multiple Time Scales

Title	Representing Human Movement in Dyadic Actions over Multiple Time Scales
Type	Research program
Question of interest	Whether the same underlying machine learning framework can be used to represent movement in dyadic actions for prediction of properties over multiple time scales.
Leaders	KTH
Other ENTIMEMENT groups involved	
Experiment type (see WP2)	○ Task2.1: Prediction in Action execution and observation ● Task2.2: Prediction in Dyadic Action execution and observation ○ Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	The objective is to study representations of movement in dyadic actions that are agnostic to the time scales of the properties to be predicted, which allows the same representation to be used for properties at different time scales. The representations are to be tested for analysis of one-on-one basketball with two players interacting with each other.
Theoretical hypotheses	Movement can be represented in a multi-scale fashion over time. Such a representation can be used to infer properties without a preselected time

	scale for prediction. In one-on-one basketball, the representation can be used to predict both immediate next actions, short-time intentions, and overall skill levels.
Operational hypotheses	Nested or stacked LSTM (long short-term memory) networks can represent movement over various time scales in parallel. Such networks can then be combined to include multiple actors and be used to infer properties that depend on all actors, such as the interplay between basketball players.
Relationship with the objectives of the project	This study relates to Task 3.4: short-term gesture prediction and Task 3.5: prediction at multiple time scales. It will explore movement analysis and prediction between multiple agents over multiple time scales.
Time schedule	Data collection, method development and analysis will be completed in the ENTIMEMENT project.
Methods	TBA
Participants	TBA
Materials	One-on-one basketball materials: In one-on-one basketball, one of the two players is the defender, and the other is the attacker. By analyzing the movements between the two players and the state of the ball, over short-term horizons, we could predict the player's movements, and over long-term horizons, player styles and the results of the battle.
Data format	RGB video and IMU data
Experimental protocol/procedure	TBA
Measures	Motion Capture: 3. Player and ball positions recorded by video cameras 4. Full body movements recorded using IMU suits
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.2.5 New-generation of radar sensors to detect mid-layer expressive gestures

Title	New-generation of radar sensors to detect mid-layer expressive gestures
Type	Research Program
Question of interest	Explore the feasibility of a new radar-based technology for motion capture analysis
Leaders	IIT-FE, UNIGEn UM-EuroMov
Other ENTIMEMENT groups involved	None
Experiment type (see WP2)	○ Task2.1: Prediction in Action execution and observation ● Task2.2: Prediction in Dyadic Action execution and observation

	● Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	○ are there relationships between the two mocap technologies? ○ Is there relevant new information in the radar technology that is complementary to the one of classical optoelectronic one?
Theoretical hypotheses	Classical mocap is very accurate in time and space. However, capturing higher-level features require a significant amount of work and yet no satisfactory solution to extract expressive features. The hypothesis is that these mid-layer features are best captured by technologies considering the body movement as whole rather that a set discrete segments moving in space.
Operational hypotheses	Radar sensors (SR) are low-power and low-complexity solution for accurate detection and tracking of moving targets. Recently, ultra-wideband (UWB) SR have gained interest owing to their ability to resolve multipaths and penetrate obstacles. It has been shown that UWB SRs can provide submeter tracking accuracy even in harsh indoor environments. Based on this fact, we will record, side by side, SR data and classical motion capture data in scenarios that are relevant for the project.
Relationship with the objectives of the project	This task will allow us to verify the potential of a whole new technology to extract complementary movement info on a different time and spatial scale.
Time schedule	Start of tests: M10
Methods	Multimodal recording of SR and mocap
Participants	At least 10 couples
Materials	SR and mocap
Data format	Matlab data structure.
Experimental protocol/procedure	Couples will have to pass each other objects of the same size but different weight. They will not know the weight in advance. In a second condition, they will be asked to pass the same objects by acting out different emotions (e.g. happiness, sadness etc.)
Measures	We will record both data set and will test whether SR can differentiate passing actions depending on weight of the object or the emotion.
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.2.6 Duomotion (Duo-Emotion)

Title	Duomotion (Duo-Emotion)
Type	Research Program
Question of interest	Several studies have focused on dyadic synchronization. Most of them have shown what are the biomechanics sources of synchronization. However psychological aspects also need to be taken into account in the motor interaction. For instance, if one partner is sad or happy it is possible that i) the quality of the synchronization would be impacted and ii) the IMS of the dyad temporarily changed at multiple time scales. Finally, iii) GMS of each emotion could be revealed.
Leaders	**UM-EuroMov**
Other EnTimeMent groups involved	UNIGE ; IIT,
Experiment type (see WP2)	○ Task2.1: Prediction in Action execution and observation ● Task2.2: Prediction in Dyadic Action execution and observation ● Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ○ None of the above
Referent scenario	EnTimeMent Dancing with Time
Research objectives	1. Design dyadic synchronization experiments to manipulate emotional and other psychological qualities among participants in motor interaction 2. Design techniques to analyse the impact of emotion in IMS and GMS 3. Design techniques to analyse multiple time scales for different motor and psychological aspects
Theoretical hypotheses	In any motor interactions psychological aspects (like emotion) are often forgotten. Emotion, personality traits, motivation… are sources of shaping the characteristics of dyadic synchronization. The hypotheses are 1) that different emotions modify partners' IMS. Positive emotions could enhance the empathy within the dyad. The two IMS would then be closer. On the other hand, negative emotions would separate the two IMS. 2) Same emotions would bring together different IMS so that a GMS of sadness, or a GMS of happiness for instance would raise. 3) The multiple time scales analysis would show different qualities of IMS at different temporal scales.
Operational hypotheses	IMS can be quantified using the similarity space (Slowinski et al., 2016), with incorporation of intentional and emotional manipulations. GMS will be under the influence of emotional differences between IMS, following the prediction that an optimum level of similarity

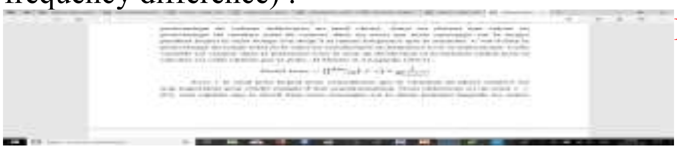
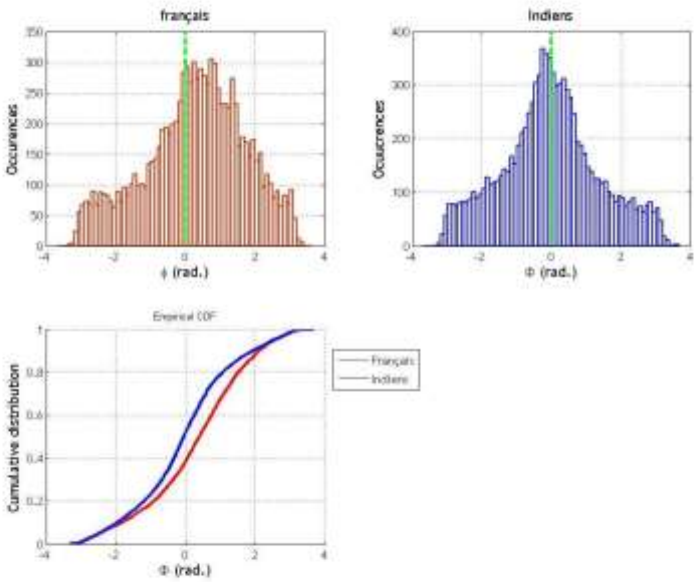
	(proximity in the similarity space) will favour the formation of a GMS, and synchronized performance. In addition, it is hypothesized that movements embedded with emotion should exist across different temporal scales. Scale-space techniques can be used to address motion segmentation and dyadic motion synchronization.
Relationship with the objectives of the project	Duomotion is part of WP2 and will lead to scenario 3.
Time schedule	Finalize protocols with partners in July 2019 (JAM meeting) Hiring of the Duomotion PhD student in September 2019 Finalizing techniques and data recording end of 2019 – beginning 2020 Multimodal recording of IMS and GMS at UM-EuroMov spring 2020 Complementary Mutimodal recordings at UNIGE in spring-summer 2020
Methods	Participants will be facing pre-recorded video of actors improvising upper-arm movements under different emotional states. The participants would have to improvise front of the actor on that video.
Participants	20 participants and 2 actors (male and female)
Materials	Large screen and pre-recorded video. Motion capture through Vicon system
Data format	Synchronized movements from video and upper-arm makers. Questionnaires of emotional state before and after each condition
Experimental protocol/procedure	Different kinds of emotion (sadness, anger, happiness, fear, disgust, neutral) will be exposed on video. Participants will improvise along with the video displayed.
Measures	Measures of frequency and phase synchronization of the dyads. Use of artificial intelligence techniques to extract and refine IMS and explore whether there are GMS based on emotion induced. Measures of the emotional state of each participant and comparison of the impact of emotion on synchronization and IMS.
Results	See Hypotheses
Descriptive results	Time series, box plots, histograms
Inference statistics	Parametric and non-parametric mixed models
Additional results	TBA
Discussion	Results will be discussed in in terms of: ● Emotional effect ● Similarity effect ● Unintended synchronization effect

2.2.7 The various Fast and slow of synchronization: A dynamical model and cultural comparison approach

Title	The various Fast and slow of synchronization: A dynamical model and cultural comparison approach
Type	Research Program
Question of interest	Development and learning in interaction with the environment, including repeated exposure and interaction with patterns determined by culture, constitute an example of very slow changes, on an individual's

	lifespan scale, that influence rhythmic skills (Jacoby & McDermott, 2017). Along this line of thinking, we aim at analysing how culture pervades across general rhythm skills and specifically determine elementary synchronization. Our first entry point was the comparison of Indian and French participants. Data collected this spring, including 15 French and 15 Indian participants, show interesting differences in the way to synchronize to a simple beat (Lagarde et al., in preparation). The data collected points at analysing further in follow ups two time scales of adaptation: Frequency and phase. For definitions and analysis, the approach uses the theoretical framework of coordination dynamics. The basic model is a non-linear model of a self-sustained oscillator (l.h.s.), forced by a periodic function and random noise (r.h.s.): $\ddot{x} + \dot{x}^3 \text{-} \dot{x} + \dot{x}.x^2 + \omega 0x = \varepsilon.\sin(\omega.t) + \sqrt{Q}.\xi t$ $\ddot{x} + \dot{x}^3 \text{-} \dot{x} + \dot{x}.x^2 + \omega 0x = \varepsilon.\sin(\omega.t) + \sqrt{Q}.\xi t$ *Eq. 1* It is well known that this model of synchronization obeys the so- called theory of Arnold's tongues (Kelso & DeGuzman, 1988), enabling identifying a priori the determiners of synchronization. From this equation relative phase dynamics can be obtained, bistable dynamics of two stable attractors, synchronization and syncopation, resp. in phase and antiphase (Kelso et al., 1990; Eq. 2): $\dot{\phi} = \Delta\omega + asin\phi - \mathrm{bsin}2\phi + \sqrt{Q}.\xi t$ $\dot{\phi} = \Delta\omega + asin\phi - \mathrm{bsin}2\phi + \sqrt{Q}.\xi t$ *Eq.2* Here we study exclusively synchronization, therefore the bistable equation Eq. 2 can be linearized to obtain further meaningful observables. We ran a first experiment (see below), and plan a follow-up examining the hypothesis that the behavioural difference observed between the Indians and French synchronization comes from sensorimotor adjustments evolving at two time scales, corresponding in short to period or phase adjustments. We aim at i) making this assumption more explicit based on available modelling, and ii) testing explicit predictions from the theory, iii) isolate essential aspects of cultural factors that determine those differences.
Leaders	Euromov-UM
Other EnTimeMent groups involved	**DU** h
Experiment type (see WP2)	○ Task2.1: Prediction in Action execution and observation ● Task2.2: Prediction in Dyadic Action execution and observation ● Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times

	○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ○ None of the above
Referent scenario	Basic single finger tapping or in follow ups index free oscillation, or vocal, sensorimotor synchronization to a beat. Using phasic stimuli perturbations to probe time scales of dynamics, that is, relaxation to attractors. The evolution of the scenario may involve using stimuli abstracted from rhythmic patterns typical of Indian music to investigate western participants synchronization to such structures. The latter calls naturally for a collaboration to identify the proper rhythmical patterns.
Research objectives	Understanding the two time scales of simple synchronization, to seek a dynamical modelling of sensorimotor synchronization considering at least two time scales instead of a unique time scale, as currently the case in the classical modelling. Furthermore, a better account of timing function in humans by encompassing cultural variants and invariants.
Theoretical hypotheses	There are several ways to achieve sensorimotor synchronization, and cultural comparison can provide further evidence of this variety, with consequences onto modelling and neuroscience assumptions. The first study was exploratory, differences in global capacity of synchronization were sought, as difference in the way synchronization was performed, considering the parameters known to determine this capacity (Arnold tongues theory). We sought to design a battery of complementary tests to estimate such dynamical parameters on an individual basis.
Operational hypotheses	A difference in global synchronization capacity, indexed by its maximal rate limits. The relative roles of the key parameters determining the quality of elementary synchronization differ between French and Indian participants.
Relationship with the objectives of the project	Contributing to the understanding of the role of multiple time scales in sensorimotor synchronization.
Time schedule	Started in march, the new data collection is planned for this fall.
Methods	In the first experiment, the task was to synchronize tapping to a periodic sound beat. The frequency of the beat was increased by .3Hz, in plateaus every 15 beats, from 1 to 6.1Hz. Complementary tests were performed to estimate individual's parameters in the frame of Arnold's tongues theory. A second experiment is planned which will consist in a similar synchronizing task, this time with constant pacing frequency and random phasic perturbation of stimuli onset. Additionally, a group of participants with a higher level of musical experience, in Indians and French participants, in their respective local music domains, will be included. Inclusion of participants from other cultures is envisioned.
Participants	For the first experiment Indians and French participants (N = 15 in each group, 11 men and 4 women, age 22 to 45), all right handed, recruited in Montpellier, were matched in pairs to control for education, age, and musical, or dance, or sports experience. Indians recruited had left India less than 2 years before the experiment, their mother tongue was Indian, their second language English, and they were not fluent in French.

Materials	A goniometer was used to collect the index finger position (metacarpophalangeal angle), connected to an A to D card, also used to collect stimuli. A second PC and the sound D to A card was used to display the stimuli.
Data format	.text files exclusively
Experimental protocol/procedure	The task was to synchronize as best as possible a tap on the table of the index finger with a sound. 3 trials were completed. The frequency of the beat was increased every 15 stimuli by 0.3 Hz. The range of the pacing frequency went from 1 to 6.1 Hz.
Measures	The relative phase between position and beats was estimated. Stationary and transients (beginning of each plateau) were separately analysed. The angular mean and dispersion were estimated. The time derivative of the relative phase was used to estimate the frequency difference between movement and stimuli, then to compute the total time spent synchronized (Dwell time, using a threshold epsilon for tolerance of frequency difference) :  Eq. 3
Results	There was no difference between the global synchronization capacity (dwell times) of the two groups. The maximal rates at which French and Indian participants were able to synchronize were comparable. However
Descriptive results	 Figure 1. Histograms of relative phases for all the plateaus for French and Indian participants (N = 9720 values; bin size 0.1 radians). The lower panel shows the cumulative distributions; a Kolmogorov-Smirnov test on the maximal difference between cumulative distributions

	confirms a significant difference between the distributions of the two groups.
Inference statistics	Non parametric, including permutation procedures, on distributions of relative phase, dwell times, estimate of individual's dynamical parameters, and relations among those parameters and synchronization performances.
Additional results	The analysis of the relation between of individual's parameters and global synchronization capacity is under way
Discussion	Results will be discussed in in terms of: ○ Influence of cultural origin onto relative phase dynamics, including stability and capacity limits. ○ Relative roles of determiners of synchronization predicted by the theory. ○ Differences in individual's dynamical parameters.

2.2.8 Understanding Entrainment Timescales During Physical Activity

Title	Understanding Entrainment Timescales During Physical Activity
Type	Research program
Question of interest	To what extent does a person's movement behaviour change during the performance of exercises based on a present other's (e.g. physio, or instructor) performance of the same exercise at the same time or his/her verbal instructions/feedback?, To what extent can two musical sonifications generated by these two be synchronised to encourage improved movement quality through sonically-supported entrainment?
Leaders	UCL
Other EnTimeMent groups involved	None
Experiment Type (see WP2)	☐ Task 2.1: Prediction in Action execution and observation ☒ Task 2.2: Prediction in Dyadic Action execution and observation ☐ Task 2.3: Prediction in Complex Action execution and observation ☐ None of the above
Use Case Scenario (see WP4)	☐ Task 4.1: Scenario 1 - Healing with multiple times ☐ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ☐ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ☒ None of the above
Research objectives	● to understand how and at what time scales entrainment may occur during dyadic physical activity ● to understand if sonification can induce/promote entrainment
Theoretical hypotheses	NA
Operational hypotheses	NA
Relationship with the objectives of the project	Aims to contribute understanding of entrainment in the context of physical activity performance

Time schedule	from December 2020
Methods	Data Collection; Analysis
Participants	People with Chronic Pain, Healthy People
Materials	Notch sensor kit, possibly Empatica sensor, video cameras and tripods, self-report materials, analysis software
Data format	None
Experimental protocol/procedure	● develop sonification ● collect data ● analysis data
Measures	TBA
Results	TBA
Descriptive results	TBA
Inference statistics	TBA
Additional results	TBA
Discussion	TBA

2.3 Prediction in Complex Action execution and observation

2.3.1 Orchestra violin sections and conductor

Title	Orchestra violin sections and conductor
Type	Experiment
Question of interest	Role of visual communication in shaping network dynamics across musicians and conductors
Leaders	IIT-FE -UNIGE
Other ENTIMEMENT groups involved	None
Experiment type (see WP2)	○ Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ● Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	The main objective is to study non-verbal communication among experts in sensori-motor synchronization such as orchestra musicians. Measures of synchronization and leadership.

Theoretical hypotheses	Movement kinematics can be used to extract the dynamical pattern of communication among orchestra players and conductors
Operational hypotheses	Acceleration profiles of different body parts movements can be used to compute causal influences (Granger analysis), among musicians and from conductor to musicians.
Relationship with the objectives of the project	This experiment will test the possibility that sensorimotor communication flows during complex multi-agent interaction along different channels of communication, at different time scales.
Time schedule	Multimodal data recordings with orchestra of Music Conservatory of Genoa and 3 different conductors at Casa Paganini was conducted during the project SIEMPRE. Data analysis was completed in the ENTIMEMENT project.
Methods	
Participants	3 conductors, 8 violinists and 10 instrumentist
Materials	Music materials: Ouverture of "Signor Bruschino", Rossini Vivaldiana, terzo movimento, Malipiero
Data format	SIEMPRE multimodal platform data
Experimental protocol/procedure	The three conductors and the orchestra executed the two pieces in a standard and two additional experimental conditions. The standard condition consisted in a normal orchestra scenario with musicians placed in a conventional spatial position. The two other conditions consisted in playing the pieces with the first violin (first row) section facing the second section (second row) thus avoiding eye contact with the conductor.
Measures	Motion capture : - violinists' bow and head position - conductors's head, left hand and baton
Results	We described the network of sensorimotor communication along two different channels of communication. The first based on instrumental movements (arm) and the other based on ancillary movements (head). Each of them was differently affected by the perturbation and thus empirically demonstrating their independence.
Descriptive results	See: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6458170/
Inference statistics	See: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6458170/
Additional results	See: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6458170/
Discussion	See: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6458170/

2.3.2 Dancing with Sync

Title	Dancing with Sync
Type	Experiment
Question of interest	In Dancing with sync, the existence of signatures of dancing expertise during voluntary group synchronization will be evaluated in a laboratory

	context, as a pre-requisite to capture the ability of dancers to maintain voluntary synchronization despite transient loss of perceptual contact
Leaders	UM-EuroMov
Other EnTimeMent groups involved	
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ● Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ○ None of the above
Referent scenario	EntiMement in dancing with Times
Research objectives	1. Test a new pendulum-based apparatus recently developed in the EnTimMent context to manipulate various qualities of group synchronization patterns at multiple temporal scales: individual characteristics including dancing expertise, type and duration of perceptual coupling, social memory, spatial organization 2. Develop specific metrics to precisely capture group synchronization regimes 3. Evaluate effect of dancing expertise, social memory, spatial configuration, and loss of perceptual coupling on synchronization regimes
Theoretical hypotheses	Expertise across multiple temporal scales related to learning (from novices to experts) modulates perceptuo-motor group synchronization
Operational hypotheses	Experts reaches group synchronization faster, maintain synchronization during loss of perceptual contact longer, and are less affected by changing spatial organization then novices
Relationship with the objectives of the project	Dancing with sync is at the intersection between WP1 (theoretical models) and WP2 (experiments) and will lead to scenario 3.
Time schedule	Develop apparatus, method, variables, analyses, and protocols in spring 2019 Data recording in Spring and Summer 2019 with non-dancers and dancers with master students enrolled on the project Dissemination in Fall 2019
Methods	Dancing in Sync will cover two experiments. In Experiment 1, 7 seated participants in different topologies (graphs) will synchronize pendulums oscillating at various similar or dissimilar frequencies. In Experiment 2, similarity will be tested with different groups of experts.
Participants	2 groups of 7 participants (Experiment 1) and 4 groups of 7 participants (2 groups of novices, 2 groups of experts)

Materials	7 pendulums with adaptable oscillating frequency (mass and mass distribution), with optical encoders for  type 1 experiments.
Data format	Synchronized analogue signals from potentiometers for type 1 experiment
Experimental protocol/procedure	Experiment 1. Three manipulations will be introduced: Topologies (complete, path, ring, star graphs), frequency similarity (homogenous, identical, different), and perceptual coupling (present, temporarily absent)
Measures	Measures of frequency and phase synchronization, at group and dyadic levels, individual contribution to group synchronization, leadership measures, use of artificial intelligence techniques to extract and refine IMS and GMS.
Results	See Hypotheses
Descriptive results	Time series, box plots, histograms
Inference statistics	Parametric and non-parametric mixed models
Additional results	TBA
Discussion	Results will be discussed in in terms of: ● Similarity effect ● Expertise effect ● Topology effect ● Perceptual effect

2.3.3 Time-to-Sync

Title	Time-to-Sync
Type	Research Program
Question of interest	In Time-to-Sync, the existence of multiple channels of perceptuomotor communication will be explored during natural and laboratory-based group synchronisation situations. Individual Motor Signatures (IMS) and group signatures (GMS) will be evaluated and modelled, and their dynamics at multiple time scales will be investigated to capture affective, emotional, and intentional qualities.
Leaders	**UM-**EuroMov (Benoît Bardy)
Other EnTimeMent groups involved	UNIGE ; IIT; WSU
Experiment type (see WP2)	● Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ● Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times

	○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Referent scenario	EntiMement in dancing with Times
Research objectives	1. Design group synchronization experiments to manipulate intentional and emotional qualities among participants 2. Develop techniques to identify IMS and GMS 3. Develop AI-based techniques to recognize intentional and emotional qualities during group interaction.
Theoretical hypotheses	Perceptuomotor group synchronisation is an essential feature of human activities. Examples include hands clapping in an audience, walking in a crowd, music playing, sport and dance. Achieving synchronisation in the group involves shared intention and perceptual interaction, but also depend on how individual motor signatures (IMS) — specific blueprints of human individuals — are assembled together to form a specific group motor signature (GMS). Theoretical hypotheses are that (i) IMS and GMS incorporate spontaneous intentional and emotional qualities, that (ii) assembling participants with different IMS affect GMS and group sensori-motor stability and performance, and that (iii) aforementioned qualities exist at different, and/or across, temporal scales.
Operational hypotheses	IMS can be quantified using the similarity space (Slowinski et al., 2016), with incorporation of intentional and emotional manipulations. GMS will be under the influence of emotional differences between IMS, following the prediction that an optimum level of similarity (proximity in the similarity space) will favour the formation of a GMS, and synchronized performance. In addition, it is hypothesized that gesture qualities (emotional and intentional components) will exist across different temporal scales
Relationship with the objectives of the project	Time-to-sync is part of WP2 and will lead to scenario 3.
Time schedule	Finalize protocols with partners in July 2019 (JAM meeting) Hiring of the Time-to-Sync PhD student in September 2019 Finalizing techniques and data recording end of 2019 – beginning 2020 Multimodal recording of IMS and GMS at UM-EuroMov spring 2020 Complementary Mutimodal recordings at UNIGE in spring-summer 2020
Methods	Time-to-Sync will involve two complementary type of experiments. Type 1 will involve participants in different topologies (graphs) to synchronize pendulums oscillating at various frequencies. Type 2 will involve participants also in different topologies synchronizing part of their body (e.g., head or arm) in more naturalistic circumstances. 
Participants	Multiple groups of 7 participants

Materials	7 pendulums with adaptable oscillating frequency (mass and mass distribution), with optical encoders for type 1 experiments. Large mocap room (NEXUS VICON and Qualisys, extended for multimodal recordings) for Type 2 experiments
Data format	Synchronized analogue signals from potentiometers for type 1 experiment Mocap and multimodal synchronized data for type 2 experiments
Experimental protocol/procedure	Three manipulations will be introduced for each type: Topologies (complete, path, ring, star graphs), intention (voluntary synchronisation vs. spontaneous synchronization) and emotion (e.g. 6 categorical emotion model).
Measures	Measures of frequency and phase synchronization, at group and dyadic levels, individual contribution to group synchronization, leadership measures, use of artificial intelligence techniques to extract and refine IMS and GMS.
Results	See Hypotheses
Descriptive results	Time series, box plots, histograms
Inference statistics	Parametric and non-parametric mixed models
Additional results	TBA
Discussion	Results will be discussed in in terms of: ● IMS and contribution to GMS ● Signatures of emotions across temporal scales ● Signatures of intention across temporal scales

2.3.4 Multiscale motor signatures in individual and joint music performance

Title	Multiscale motor signatures in individual and joint music performance
Type	Research program
Question of interest	Role of similarity in motor signatures at multiple timescales in determining compatibility of action styles in musical performers
Leaders	UNIGE; UM-EuroMov; DU; WSU
Other ENTIMEMENT groups involved	
Experiment type (see WP2)	○ Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ● Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	To develop techniques for analysing motor signatures from musicians' movement kinematics at multiple timescales and to investigate the role

	of similarity in these signatures in determining compatibility in action style during joint musical performance
Theoretical hypotheses	Similarity in motor signatures at multiple timescales will determine the quality of interpersonal coordination during joint music performance by enhancing compatibility in action style
Operational hypotheses	Measures of motor signatures based on multi-dimensional scaling techniques applied to movement velocities for different body segments moving at different timescales (e.g., arm movement vs body sway) will explain variance in coordination across instrument duos.
Relationship with the objectives of the project	This series of experiments will investigate how information at multiple timescales explains predictive processes in complex joint action execution and observation in terms of compatibility of motor signatures and action styles.
Time schedule	Develop analytical techniques with existing data from the TELMI corpus of violin performances and other existing datasets, and in parallel build a repository of multimodal recordings of group musical performance.
Methods	Motor signature analysis and exploring with machine learning techniques Synchronization techniques Multimodal recording with motion capture, audio, video, EMG, and respiration.
Participants	Expert violin performers and possible other instrumentalists; Musicologists for the selection of music fragments used in the experiment; observers for perceptual studies
Materials	Music materials: From TELMI corpus and possible ethnomusicological corpus. Duo and small ensemble musical pieces, including newly composed pieces designed to elicit particular kinds of interaction between performers.
Data format	SIEMPRE multimodal platform data
Experimental protocol/procedure	Motor signatures will be analysed based on movement velocities for different body segments moving at different timescales (e.g., arm movement vs body sway) using multi-dimensional scaling techniques. Machine learning techniques will be employed to explore the relationship between the motor signatures at multiple timescales. Objective measures of interpersonal coordination in joint music performance will be computed in multiple modalities (e.g., audio, video, mocap) and at different timescales. Subjective measures of coordination based on observer perceptual judgements will be collected. Measures of multi-timescale motor signature similarity will be used to predict objective and subjective measures of coordination.
Measures	Motion capture, video, audio, EMG, respiration, observer perceptual judgements
Results	Consolidation of techniques and implementation of software modules, which can be used in project Scenarios
Descriptive results	TBA
Inference statistics	TBA

Additional results	TBA
Discussion	TBA

2.3.5 Tracking the leader: gaze behaviour in group interactions

Title	Tracking the leader: gaze behaviour in group interactions
Type	Experiment
Question of interest	Can social gaze behaviour reveal the leader during real-world group interactions?
Leaders	IIT
Other ENTIMEMENT groups involved	None
Experiment type (see WP2)	○ Task2.1: Prediction in Action execution and observation ○ Task2.2: Prediction in Dyadic Action execution and observation ● Task2.3: Prediction in Complex Action execution and observation ○ None of the above
Use Case scenario (see WP4)	○ Task 4.1: Scenario 1 - Healing with multiple times ○ Task 4.2: Scenario 2 - Chronic musculoskeletal pain management with multiple times ○ Task 4.3: Scenario 3 - EntimeMent in dancing with Times ● None of the above
Research objectives	Stereotypical thinking links leadership to prolonged gazing towards leaders (Hall et al., 2005) and longer mutual gazing in response to interactions initiated by leaders (Carney et al., 2005). However, evidence for an actual relationship between leadership and social gaze behaviours is limited. To date, investigations on the influence of leadership on gaze behaviour have focused on computer-based paradigms that do not provide any opportunity for social interaction (Capozzi and Ristic, 2018; Koski et al., 2015; Risko et al., 2016). The aim of the present study was to develop a novel approach to investigate how leadership shapes gaze dynamics during real-world human group interactions.
Theoretical hypotheses	Multi-party gaze features code implicit semantics of social gaze behaviours, and more specifically, leadership.
Operational hypotheses	The basic idea for establishing a relationship between social gaze behaviour and leadership was to conceptualize multi-party gaze features as patterns and to treat the analysis as a pattern classification problem: can a classifier applied to the visual behaviour pattern of real people interacting in small groups reveal the leader?
Relationship with the objectives of the project	Test social gaze behaviour can reveal the leader during real-world group interactions.
Time schedule	Multimodal data recordings completed before project start. Data analysis was completed in the ENTIMEMENT project
Methods	
Participants	16 groups composed of four previously unacquainted individuals
Materials	Each group of participants was asked to complete one of two versions of a survival task (“Winter” or “Desert”; Johnson and Johnson, 1994). The

	task involved rank-ordering 12 ordinary items (e.g., a map, a mirror, a chocolate bar) based on their utility for group-surviving in a hostile environment. The use of pen paper was not allowed.
Data format	https://ars.els-cdn.com/content/image/1-s2.0-S2589004219301725-mmc2.xlsx
Experimental protocol/procedure	Participants were assigned to one of four-person groups, for a total of sixteen groups. Eight participants classified as leaders with a democratic leadership style and eight participants classified as leaders with an autocratic leadership style were randomly assigned as 'designated leaders' to one of the sixteen groups. Forty-eight of the potential followers were also randomly assigned to each group. Each group of participants was asked to complete one of two versions of a survival task (see materials). Democratic leadership is expected to be more effective under situational conditions of low time-pressure, whereas autocratic leaderships is expected to be more effective under situational conditions of high time-pressure (Fiedler, 2006; Pierro et al., 2003). To manipulate situational conditions, a time-pressure manipulation was applied (Chirumbolo et al., 2004; De Grada et al., 1999; Kruglanski and Freund, 1983; Pierro et al., 2003). Groups assigned to the high time-pressure situation (n = 8) were instructed to perform the assigned task as quickly as possible, with a clear instruction that time was a critical demand to their task. Groups assigned to the low timepressure situation (n =8) were instead encouraged to take their time to reach a decision with no specific time demand. The orthogonal manipulation of leadership styles and situational conditions resulted in two high-fit conditions (Democratic - Low time-pressure, Autocratic - High time-pressure) and two low-fit conditions (Democratic - High time-pressure, Autocratic - Low time-pressure) (Figure 1 A; see also SI and Figure S1 for group composition and manipulation checks).
Measures	Four AXISP1346 multi-view streaming cameras (1280x1024 pixels resolution, 20 frame per second frame rate) were used for individual video recording of the upper part of the body (head and shoulders) of each group member. Individual videos were used for VFOA modelling and visual behaviour features extraction.
Results	We found that social gaze behaviour distinctively identified group leaders. Crucially, the relationship between leadership and gaze behaviour generalized across democratic and autocratic leadership styles under conditions of low and high time-pressure, suggesting that gaze can serve as a *general marker of leadership*. These findings provide the first direct evidence that group visual patterns can reveal leadership across different social behaviours and validate a new promising method for monitoring natural group interactions.
Descriptive results	https://www.sciencedirect.com/science/article/pii/S2589004219301725?via%3Dihub
Inference statistics	https://www.sciencedirect.com/science/article/pii/S2589004219301725?via%3Dihub

Additional results	https://www.sciencedirect.com/science/article/pii/S2589004219301725?via%3Dihub
Discussion	https://www.sciencedirect.com/science/article/pii/S2589004219301725?via%3Dihub

3. HUMAN MOVEMENT DATA-SETS

This section includes tables describing pre-existing as well as early versions of the movement datasets developed by EnTimeMent partners. Appendix 1 provides a survey of third-party movement datasets.

3.1.1 Lightness and Fragility IMU and video dataset

Title	Lightness and Fragility IMU and video dataset
Type	IMU and video
Question of interest	Investigate movement qualitites
Owner	UNIGE
Other ENTIMEMENT groups involved	UM, freely available to the EnTimeMent consortium and the research community
Participants	12 dancers
Short description and objective	The dataset consists of 120 segments of synchronized video and IMU data. Each segments has duration of about 10s. Videos are full-body, with blurred faces of the dancers, to focus only on body movement (without facial expression). IMU sensors with 9dof each are placed on wrists, ankles, and coccyx.
Kind of data	Text and video files with SMPTE timecode. Video files have the SMPTE code in one of the audio channels.
Sensors	XOSC IMUs and videocameras
Privacy status	Freely available to the research community
Data format	Text and mp4 video
Link	http://beatricedegelder.com/documents/Vaessen2018.pdf

3.1.2 TELMI Violin Performance Dataset

Title	TELMI Violin Performance Dataset
Type	Mocap, Video, Kinect, Audio and MYO
Question of interest	Investigation of movement in violin performance, quality of the performance
Owner	UNIGE
Other ENTIMEMENT groups involved	Freely available to the EnTimeMent consortium and the research community
Participants	
Short description and objective	The dataset consists in multimodal recordings of 4 professional violinist from Royal College of Music of London performing 41 exercises from classical pedagogy repertoire, collecting in the TELMI Multimodal

	Archive. Furthermore, the dataset includes recordings of three students and one teacher performing a programme of 18 exercises from the original list of 41.
Kind of data	Mocap, audio, Kinect, EMG data and video files with SMPTE timecode. Video files have the SMPTE code in one of the audio channels.
Sensors	13-cameras Qualysis motion capture system, cameras, MYO sensors, Kinect
Privacy status	Freely available for the research community
Data format	.tsv, .qtm, .mp4, .aif, .txt
Link	TELMI Archive paper

3.1.3 UNIGE EnTimeMent Multimodal Recordings Dataset

Title	UNIGE EnTimeMent Multimodal Recordings Dataset
Type	Qualisys Motion Capture data, synchronized via SMPTE with multichannel audio (including audio respiration), multiple professional videocameras, IMUs, EMG and possible other biometric data.
Question of interest	Investigation of prediction and analysis at multiple temporal scales of individual as well as group behaviour.
Owner	UNIGE
Other ENTIMEMENT groups involved	Consortium
Participants	Healthy adults and children
Short description and objective	Provide the necessary dataset recordings for several experiments in EnTimeMent (see previous section)
Kind of data	Mocap, audio, Kinect, EMG, IMU
Sensors	Qualysis motion capture cameras, videocameras, microphones, MYO, Kinect, IMU, XOSC and other sensors
Privacy status	Freely available to all consortium partners.
Data format	.tsv, .qtm, .mp4, .aif, .txt (IMU and EMG)
Link	https://entimement.dibris.unige.it/user_files/CPIM-ETM-LabelsListForRecordings.PDF

3.1.4 UCL Emo-Pain dataset

Title	EmoPain
Type	Motion capture, surface electromyography
Question of interest	Movement behaviour in people with chronic pain
Owner	UCL
Other ENTIMEMENT groups involved	None
Participants	People with chronic low back pain and healthy people

Short description and objective	The data was captured from participants while they performed physical exercises typically prescribed for chronic pain physical rehabilitation, and similar to everyday movements (sit-to-stand-to-sit, standing on one leg, forward reaching, bending, walking, sitting, standing)
Kind of data	Body movement data
Sensors	Full-body gyroscope sensors, surface electromyography
Privacy status	Anonymised data available to consortium partners on request, following GDPR and UCL research ethics restrictions
Data format	mat files
Link	Not publicly available

3.1.5 IEMP Data Collection

Title	Interpersonal Entrainment in Music Performance (IEMP) Data Collection
Type	Audio, video and annotation data of musical performances in diverse genres
Question of interest	Interpersonal synchronisation and coordination in musical ensembles
Owner	DU
Other ENTIMEMENT groups involved	UNIGE, UWS
Participants	Professional and semi-professional musicians
Short description and objective	The IEMP Collection, shared publicly on Open Science Framework, contains recordings and annotations of musical performances in six genres. Contents are summarized in the table.
Kind of data	Audio, Video, and Time-stamped text annotations: musical structure, metre, event onsets, onsets assigned to metrical positions, movement extracted using Optical Flow algorithm in Eyesweb (part only). Code also shared, linked under Technical Resources.
Sensors	Digital audio and video recorders
Privacy status	Publicly shared. Restrictions on non-research (inc. commercial) re-use.
Data format	WAV, MP4, CSV, TXT
Link	https://osf.io/37fws/

IEMP Data Collection Contents

Genre	Abbr.	Origin	Group size	Instrumentation	Size of corpus	Dur. (min)	Researcher

North Indian Raga	NIR	North India	2-6	Sitar, sarod or guitar + tabla or vocal, harmonium + tabla (tanpura drone not analysed)	8 raga performances pieces, Mean duration = 3,000 seconds (*SD* = 582)	413	M. Clayton, L. Leante
Uruguayan Candombe	UC	Uruguay	3-4	Chico, piano and repique drums	12 takes, M = 175.5s (*SD* = 30.9)	35	L. Jure, M. Rocamora
Malian Jembe	MJ	Mali	2-4	Jembe and dundun drums	15 takes of 3 pieces, M = 202s (*SD* = 69.1)	51	R. Polak
Cuban Son and Salsa	CSS	Cuba	7	Bass, Spanish guitar, tres, clave, bongos and other percussion, trumpet, vocals	5 songs, M = 398s (*SD* = 45.5)	33	A Poole
Tunisian Stambeli	TS	Tunisia	≥4, 2 parts analysed	Gumbri (lute) + shqashiq (cymbals), vocals. Nb no video.	4 tracks comprising 8 pieces, M = 259.8s (*SD* = 105.2)	35	R. Jankowsky
String Quartet	SQ	UK	4	Violin x 2, viola, cello. Nb no video.	2 takes each of 2 movements, M = 290.2s (SD = 20.3)		
String Quartet	SQ	Europe	4	Violin x 2, viola, cello	2 takes each of 2 movements, extracts	6	M. Clayton, T. Eerola, K. Jakubowski

4. EARLY PUBLICATIONS FROM THE CONSORTIUM

Scientific publications are already available from the consortium. They are available as Open Access, and can be found collectively from the following project web page: https://entimement.dibris.unige.it/documents

○ ANNEX 1 A SURVEY OF EXISTING BODY MOVEMENT DATASETS

○ Survey Summary

Survey Leaders	UCL
Survey Exclusion Criteria	• of static pose • of face/hand/gaze only or single body location • based on top view camera only • not of humans or mainly of just objects or animals • not particularly involving movement or of sedentary activities • based on movement sensor on object rather than human
Total Number of Datasets	134
Number by Sensor Category	Based on Video only = 71; Based on Inertia sensors = 62; Based on Electromyography only = 1
Number by Dataset Availability	Data webpage published = 87; Data webpage not published = 57

○ Survey Details - only for the datasets with published webpages

The details are in two parts (Part I - Videos only, Part II - including inertia sensors with electromyography), each ordered by publication year.

Dataset Name	Dataset Author & Year	Source of Dataset	Type of Data	Data Size	Activities in Data	Data Webpage
			PART I - VIDEOS ONLY			
Large Scale Combined RGB-D Action Dataset	Zhang et al. 2018	multiple datasets	videos, depth videos and human action labels	4953 sequences covering 94 human actions	various	https://www.uow.edu.au/~wanqing/#Datasets
DeepMind Kinetics Human Action Video dataset	Kay et al. 2017	YouTube videos	videos and human action labels	500,000 video clips covering 600 human actions	everyday type	http://deepmind.com/kinetics

TenniSet	Faulkner and Dick 2017	YouTube videos	videos and event labels	787,600 video frames covering 6 tennis event types	tennis matches	https://github.com/HaydenFaulkner/Tennis
Atomic Visual Actions Dataset	Gu et al. 2017	YouTube videos	videos and human action/interaction labels	392,426 video clips covering 60 human actions/interaction types	unknown	https://research.google.com/ava/
Human Action Clips and Segments Dataset	Zhao et al. 2017	YouTube videos	videos and human action labels	1.55M video clips covering 200 human actions	various	http://hacs.csail.mit.edu/
MultiTHUMOS dataset	Yeung et al. 2017	YouTube videos	videos and human action labels	400 videos of THUMOS14 covering 65 human actions (including THUMOS14's)	various	http://ai.stanford.edu/~syyeung/everymoment.html
The “something something” video database	Goyal et al. 2017	Recorded in acted scenarios	videos (hand only)	220,847 videos covering 174 hand-object interaction types	hand-object interaction scenarios	https://20bn.com/datasets/something-something/v2
Dataset Name	**Dataset Author & Year**	**Source of Dataset**	**Type of Data**	**Data Size**	**Activities in Data**	**Data Webpage**
Daily Action Localization in YouTube	Weinzaepfel et al. 2017	YouTube videos	videos with person bounding box and human action labels	510 videos covering 10 human actions	everyday type	http://thoth.inrialpes.fr/daly/
MSR-Video to Text	Xu et al. 2016	Videos on the internet	videos	40,000 clips	everyday type	http://ms-multimedia-challenge.com/2017/dataset
NCAA Basketball Dataset	Ramanathan et al. 2016	YouTube videos	videos with player bounding box and event labels	14,548 video clips covering 11 event types	basketball games	http://basketballattention.appspot.com/#dataset
ACT dataset	Wang et al. 2016	YouTube videos	video clips and activity labels	11,234 video clips covering 43 activties	unknown	http://www.cs.cmu.edu/~xiaolonw/actioncvpr.html
Hollywood2Tubes	Mettes et al. 2016	Movies	videos and human action labels and bounding box of persons in some sections	1,707 of Hollywood2 covering 12 actions of Hollywood2	unknown	https://staff.fnwi.uva.nl/p.s.m.mettes/codedata.html

Charades	Sigurdsson et al. 2016	Recorded in acted scenarios	videos	9,848 video sequences covering 157 human actions	household actvities	https://allenai.org/plato/charades/
UWA3D Multiview Activity II Dataset	Rahmani et al. 2016	Recorded in acted scenarios	depth videos and activity labels	1,200 sequences covering 30 activities	various	http://staffhome.ecm.uwa.edu.au/~00053650/databases.html
MPII Cooking 2	Rohrbach et al. 2016	Recorded in naturalistic scenarios in lab	videos and human action labels, some also with labels of pose of anatomical segments, and some further with hand region marked	273 videos covering 87 human actions	cooking	https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/human-activity-recognition/mpii-cooking-2-dataset/
MERL Shopping Dataset	Singh et al. 2016	Recorded in naturalistic scenarios in lab	videos and human action labels	96 videos covering 5 human actions	shopping	http://www.merl.com/demos/merl-shopping-dataset
ActivityNet	Heilbron et al. 2015	Videos on the internet	videos and human action labels	19,994 videos covering 200 human activity labels	everyday type	http://activity-net.org/
MPII Movie Description Dataset	Rohrbach et al. 2015	Movies	videos with audio transcript	68,337 video clips	everyday type	https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/vision-and-language/mpii-movie-description-dataset/
Montreal Video Annotation Dataset	Torabi et al. 2015	Movies	videos with audio transcript	48,986 video clips	everyday type	https://mila.quebec/en/publications/public-datasets/m-vad/
Dataset Name	**Dataset Author & Year**	**Source of Dataset**	**Type of Data**	**Data Size**	**Activities in Data**	**Data Webpage**
THUMOS15 Challenge Dataset	Idrees et al. 2017 (dataset was published in 2015)	YouTube videos	videos and activity labels, with additional sub-action labels	18,404 videos covering 101 activities	unknown	http://www.thumos.info/download.html
Office Activity Dataset	Wang et al. 2015	Recorded in acted scenarios	videos, depth videos, and activity labels	1,180 sequences	office type activites	http://www.sysu-hcp.net/resources/

Sports-1M Dataset	Karpathy et al. 2014	YouTube videos	videos and activity labels	1,133,158 sequences covering 487 activities	sports	https://cs.stanford.edu/people/karpathy/deepvideo/
Breakfast	Kuehne et al. 2014	Recorded in naturalistic scenarios in lab	videos and human action labels	1,989 sequences covering 10 human actions	cooking	http://serre-lab.clps.brown.edu/resource/breakfast-actions-dataset/
LIRIS human activities dataset	Wolf et al. 2014	Recorded in acted scenarios	videos, depth videos and activity labels (with bounding box)	covering 10 activities	various	https://projet.liris.cnrs.fr/voir/activities-dataset/
joint-annotated HMDB	Jhuang et al. 2013	Online videos	video clips and human action labels from HMDB51, and 2D positions of full body joints of the subject	928 video clips covering 21 action categories from the HMDB51	various	http://jhmdb.is.tue.mpg.de/
Penn Action Dataset	Zhang et al. 2013	Online videos	videos and actitivity labels with label of anatomical segment involved and its bounding box	2,326 covering 15 activities	sports	https://github.com/dreamdragon/PennAction
Mivia Action Dataset	Carletti et al. 2013	Recorded in acted scenarios	depth videos and human action labels	500 sequences covering 7 human actions	various	https://mivia.unisa.it/datasets/video-analysis-datasets/mivia-action-dataset/
Osaka University Kinect Action Data Set	Mansur et al. 2013	Recorded in acted scenarios	videos, depth videos and human action labels	covering 10 human actions	sports	http://www.am.sanken.osaka-u.ac.jp/~mansur/dataset.html
DMLSmartActions dataset	Amiri et al. 2013	Recorded in acted scenarios	videos, depth videos and human action labels	932 videos covering 25 human actions	everyday type	http://dml.ece.ubc.ca/data/smartaction/
3D Action Pairs aka MSRActionPair dataset	Oreifej and Liu 2013	Recorded in acted scenarios	depth image sequences and human action labels	covering 12 human actions	everyday type	http://www.cs.ucf.edu/~oreifej/HON4D.html#New%20dataset%20-%20MSR%20Action%20Pairs
Dataset Name	**Dataset Author & Year**	**Source of Dataset**	**Type of Data**	**Data Size**	**Activities in Data**	**Data Webpage**
UCF101 - Action Recognition Data Set	Soomro et al. 2012	YouTube videos	videos and activity labels	13,320 videos covering 101 activities	sports, everyday type	https://www.crcv.ucf.edu/research/data-sets/human-actions/ucf101/

ASLAN	Kliper-Gross et al. 2012	YouTube videos	videos and human action labels	3,631 video clips covering 432 human actions	everyday type	https://talhassner.github.io/home/projects/ASLAN/ASLAN-main.html
ACT4[2]	Cheng et al. 2012	Recorded in acted scenarios	videos, depth videos and activity labels	6,844 covering 14 activities	everyday type	https://sites.google.com/site/qinleisite/Home/dataset
BIT-Interaction Dataset	Kong et al. 2012	Recored in acted scenarios	videos and human interaction labels	400 videos covering 8 human interaction scenarios	human-human interaction activities	https://sites.google.com/site/alexkongy/software
UTKinect-Action3D Dataset	Xia et al. 2012	Recorded in acted scenarios	videos, depth videos	200 sequences covering 10 human actions	everyday type	http://cvrc.ece.utexas.edu/KinectDatasets/HOJ3D.html
Depth-included Human Action video	Lin et al. 2012	Recorded in acted scenarios	videos, depth videos	483 sequences covering 23 human actions	various	http://mclab.citi.sinica.edu.tw/dataset/dha/dha.html
Zhang and colleagues 2012	Zhang et al. 2012	Recorded in acted scenarios	videos, depth videos	87 sequences covering 8 human actions	falls and movements with poses similar to falls	http://vlm1.uta.edu/~zhangzhong/fall_detection/
Actions for Cooking Eggs Dataset	Shimada et al. 2012	Recorded in naturalistic scenarios in lab	videos and depth videos (showing hands only)	25 sequences covering 8 human actions	cooking eggs	http://www.murase.m.is.nagoya-u.ac.jp/KSCGR/download.html
MPII Cooking Activities Dataset	Rohrbach et al. 2012	Recorded in naturalistic scenarios in lab	videos and human action labels, some also with labels of pose of anatomical regions	44 videos covering 65 human actions	cooking	https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/human-activity-recognition/mpii-cooking-activities-dataset/
Human Motion DataBase	Kuehne et al. 2011	Online videos	videos (full body visible only for about half of the videos and human action labels	7,000 videos covering 51 human actions (facial and bodily)	various	http://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/
VideoPose2.0	Sapp et al. 2011	Friends, Lost TV series	2 to 3 secs long video clips (not usually full body)	44 video clips	various	http://bensapp.github.io/videopose-dataset.html
VIRAT Video Dataset	Oh et al. 2011	unknown	videos and human action labels	23 human actions	everyday type	http://www.viratdata.org/

Dataset Name	Dataset Author & Year	Source of Dataset	Type of Data	Data Size	Activities in Data	Data Webpage
Olympic Sports Dataset	Niebles et al. 2010	YouTube videos	videos and human action labels	800 video sequences covering 16 human actions	sports	http://vision.stanford.edu/Datasets/OlympicSports/
TV Human Interaction Dataset	Patron-Perez et al. 2010	TV shows	video clips with upper body bounding box, and head orientation and interaction labels	300 video clips	hand shake, high five, hug, kiss	http://www.robots.ox.ac.uk/~alonso/tv_human_interactions.html
Multicamera Human Action Video Dataset	Singh et al. 2010	Recorded	video clips and human action labels	1904 video clips (only 952 is public) covering 17 human actions	various	http://velastin.dynu.com/MuHAVi-MAS/
i3DPost multi-view and 3D human action/interaction database	Gkalelis et al. 2009	Recorded in acted scenarios	videos (some face only) and human action labels	104 videos covering 12 human actions (+ emotional facial expressions only)	various	http://kahlan.eps.surrey.ac.uk/i3dpost_action/
UT-Interaction	Ryoo and Aggarwal 2009	Recored in acted scenarios	videos and human interaction labels with bounding box	20 video sequences covering 6 human interaction scenarios	various	http://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html
HOHA	Laptev et al. 2008	movies	video and human action labels	444 video sequences covering 8 human actions	everyday type	https://www.di.ens.fr/~laptev/actions/
Virtual Human Action Silhouette data	Ragheb et al. 2008	Artificially generated	videos	180 covering 20 human actions	various	http://velastin.dynu.com/VIHASI/
Weizmann Action Dataset	Gorelick et al. 2007	Recorded in acted scenarios	videos and human action labels	90 sequences covering 10 human actions	various	http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html
Inria Xmas Motion Acquisition Sequences	Weinland et al. 2006	Recorded in acted scenarios	videos, silhoutte videos and human action labels	covering 13 human actions	everyday type	http://4drepository.inrialpes.fr/public/viewgroup/6
HumanID Gait Challenge dataset	Phillips et al. 2005	Recorded in acted scenarios	videos	1870 videos	walking	http://www.eng.usf.edu/cvprg/GaitBaseline/index.html

Video Event Detection dataset	Ke et al. 2005	unknown	videos and human action labels	48 videos covering 4 human actions	everyday type	http://www.yanke.org/
KTH Human Action dataset	Schuldt et al. 2004	Recorded in acted scenarios	videos and human action labels	2,391 sequences covering 6 human actions	various	http://www.nada.kth.se/cvap/actions/
Caviar Data	Fisher 2004	Shopping mall surveillance	videos and activity labels (with bounding box of subject)	28 video sequences covering 6 activities	various	homepages.inf.ed.ac.uk/rbf/CAVIARDATA1
Dataset Name	**Dataset Author & Year**	**Source of Dataset**	**Type of Data**	**Data Size**	**Activities in Data**	**Data Webpage**
			PART II - BASED ON INERTIA SENSORS OR ELECTROMYOGRAPHY			
CMU Graphics Lab Motion Capture Database	unknown	Recorded in acted scenarios	videos and full body joints positions and activity labels	2,605 sequences	various	http://mocap.cs.cmu.edu/
UOW Online Action3D Dataset	Tang et al. 2018	Recorded in acted scenarios	videos, depth videos and 3D full body positions and human action labels	covering 20 human actions	various	https://www.uow.edu.au/~wanqing/#UOWActionDatasets
NTU RGB+D Action Dataset	Shahroudy et al. 2016	Recorded in acted scenarios	videos, depth videos and fullbody positions and human action labels	56,880 sequences covering 40 human actions	everyday type	http://rose1.ntu.edu.sg/Datasets/actionRecognition.asp
UTD Multimodal Human Action Dataset	Chen et al. 2015	Recorded in acted scenarios	videos, depth videos and fullbody positions, triaxial accelerometer, gyroscope, and magnetometer data and human action labels	861 sequences covering 27 human actions	various	http://www.utdallas.edu/~cxc123730/UTD-MHAD.html
Watch-n-Patch	Wu et al. 2015	Recorded in acted scenarios	videos, depth videos, fullbody positions and human action labels	458 videos covering 21 human actions	house and office work	http://watchnpatch.cs.cornell.edu/
Multi-modal & Multi-view & Interactive dataset	Xu et al. 2015	Recorded in acted scenarios	videos, depth videos, fullbody positions and human action labels	1760 sequences covering 22 human action categories	various	http://media.tju.edu.cn/datasets.html

G3Di	Bloom et al. 2015	Recorded in naturalistic scenarios in lab	videos, depth videos, fullbody positions and human action labels	covering 18 human actions	gaming activities	http://dipersec.king.ac.uk/G3D/index.html
ShakeFive (1 & 2)	van Gemeren et al. 2014	Recorded in acted scenarios	video and fullbody positions and activity labels	153 sequences covering 8 activities	everyday interaction type	http://www2.projects.science.uu.nl/shakefive/
UPCV Gait Dataset & UPCV Gaik K2 Dataset	don't know	Recorded in acted scenarios	positions of fullbody joints	not known	walking	http://www.upcv.upatras.gr/personal/kastaniotis/datasets.html
UPCV Action Dataset	Theodorakopoulos et al. 2014	Recorded in acted scenarios	videos, depth videos, fullbody positions, and human action labels	covering 10 human actions	various	http://www.upcv.upatras.gr/personal/kastaniotis/datasets.html
Northwestern-UCLA Multiview Action 3D Dataset	Wang et al. 2014	Recorded in acted scenarios	videos, depth videos, fullbody positions and human action labels	covering 10 human actions	various	http://users.eecs.northwestern.edu/~jwa368/my_data.html

Dataset Name	**Dataset Author & Year**	**Source of Dataset**	**Type of Data**	**Data Size**	**Activities in Data**	**Data Webpage**
UCF Kinect	Ellis et al. 2013	Recorded in acted scenarios	positions of fullbody joints	1,280 sequences covering 16 human actions	gaming actions	http://www.syedzainmasood.com/research.html
IAS-Lab Action Dataset	Munaro et al. 2013	Recorded in acted scenarios	videos, depth videos, fullbody joints positions, and human action labels	540 sequences covering 15 human actions	various	http://robotics.dei.unipd.it/actions/index.php/overview
Berkeley Multimodal Human Action Database	Ofli et al. 2013	Recorded in acted scenarios	video and fullbody positions and accelerometer and human action label	660 sequences covering 11 human actions	everyday type	http://tele-immersion.citris-uc.org/berkeley_mhad
Kinect-Based 3D Human Interaction Dataset	Hu et al. 2013	Recorded in acted scenarios	positions of fullbody joints and human interaction labels	covering 6 human interaction scenarios	human-human interaction activities	http://www.lmars.whu.edu.cn/prof_web/zhuxinyan/DataSetPublish/dataset.html

Cornel Activity Dataset-120	Koppula et al. 2013	Recorded in acted scaenarios	videos, depth videos, positions for fullbody joints with activity labels	120 sequences covering 10 activities (parent) and 10 human actions (child)	everyday type	http://pr.cs.cornell.edu/humanactivities/data.php
Florence 3D Action dataset	Seidenari et al. 2013	Recorded in acted scenarios	videos and full body positions and activity labels	215 sequences	everday type	https://www.micc.unifi.it/resources/datasets/florence-3d-actions-dataset/
Microsoft Research Cambridge-12 Kinect gesture data set	unknown	Recorded in acted scenarios	3D positions of joints and gesture labels	594 sequences covering 12 human actions	unknown	https://www.microsoft.com/en-us/download/details.aspx?id=52283&from=http%3A%2F%2Fresearch.microsoft.com%2Fen-us%2Fum%2Fcambridge%2Fprojects%2Fmsrc12%2F
Stony Brook University Kinect Interaction Dataset	Yun et al. 2012	Recorded in acted scenarios	videos, positions of fullbody joints, and activity labels	300 sequences covering 8 activities	human-human interaction activities	http://www3.cs.stonybrook.edu/~kyun/research/kinect_interaction/index.html
MSRDailyActivity3D Dataset	Wang et al. 2012	Recorded in acted scenarios	videos, depth videos, positions of fullbody joints, and action labels	320 sequences covering 16 activities	various	https://www.uow.edu.au/~wanqing/#Datasets
G3D	Bloom et al. 2012	Recorded in acted scenarios	videos, depth videos, positions of full body and activity labels	70 sequences covering 20 human actions	gaming activities	http://dipersec.king.ac.uk/G3D/
Physical Activity Monitoring for Aging People Dataset	Reiss and Stricker 2012	Recorded in acted scenarios	accelerometer, gryoscope, and magnetometer, heart rate, and activity labels	3,850,505 sequences covering 18 activities	various	http://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring
Dataset Name	**Dataset Author & Year**	**Source of Dataset**	**Type of Data**	**Data Size**	**Activities in Data**	**Data Webpage**
Cornel Activity Dataset-60	Sung et al. 2011 & Sung et al. 2012	Recorded in acted scenarios	videos, depth videos positions for fullbody joints, and activity labels	60 sequences covering 12 activities	everday type	http://pr.cs.cornell.edu/humanactivities/data.php

MSR Action3D & MSRAction3DExt Dataset	Li et al. 2010 & Wang et al. 2016	Recorded in acted scenarios	depth map and positions for joints and human action labels	567 (MSR Action3D only) & 1379 (both) sequences covering 20 human actions	various	https://www.uow.edu.au/~wanqing/#Datasets (MSR Action3D)
Daphnet Freezing of Gait Data Set	Bachlin et al. 2010	Recorded in the lab	accelerometer and freezing of gait labels	237 sequences	walking	https://archive.ics.uci.edu/ml/datasets/Daphnet+Freezing+of+Gait
Opportunity	Roggen et al. 2010	Recorded in acted scenarios	accelerometer, positions and human action labels	not known	everyday type	http://www.opportunity-project.eu/challengeDataset.html
HumanEva datasets	Sigal et al. 2010	Recorded in acted scenarios	video and fullbody positions	56 sequences covering 5 activities	walk, jog, throw/catch, box, gesturing	http://humaneva.is.tue.mpg.de
TUM Kitchen Data Set	Tenorth et al. 2009	Recorded in acted scenarios	videos and joint angles and positions (both full body) and human action labels for the different anatomical regions	not known	household actvities	https://ias.in.tum.de/dokuwiki/software/kitchen-activity-data
Carnegie Mellon University Multimodal Activity Database	de La Torre et al. 2008	Recorded in naturalistic scenarios in lab	video, accelerometer, gyroscope and magnetometer, positions of full body joints	covering 17 human actions	cooking	http://kitchen.cs.cmu.edu/
Skoda	Stiefmeier et al. 2008	Recorded in naturalistic scenarios	inertia sensor, force sensitive ressistor and activity labels	3680 sequences	car assembly quality assurance activities	http://har-dataset.org/doku.php?id=wiki:dataset
Motion Capture Database HDM05	Muller et al. 2007	Recorded in acted scenarios	fullbody positions and videos and activity labels	1457 sequences covering 100 activities	various	http://resources.mpi-inf.mpg.de/HDM05/
PACO Body Movement Library	Ma et al. 2006	Recorded in acted scenarios	positions of fullbody joints and videos and human action labels with emotional state labels	4080 sequences covering 3 human actions, and 4 emotional states	walking, knocking, lifting and throwing, with affective elements	http://paco.psy.gla.ac.uk/index.php/res/download-data

5. REFERENCES

Ahern, D.K. et al. (1988) Comparison of lumbar paravertebral EMG patterns in chronic low back pain patients and non-patient controls. *Pain*, *34*(2), 153–160.
Altun, K., et al. (2010). Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition*, *43*(10), 3605–3620. https://doi.org/10.1016/j.patcog.2010.04.019
Amiri, M. S. (n.d.), et al. (2018). Non-Intrusive Human Activity Monitoring in a Smart Home Environment.
Arn, R. et al. (2018). Motion Segmentation via Generalized Curvatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PP*(X), 1. https://doi.org/10.1109/TPAMI.2018.2869741
Bächlin, M. et al. (2010). Wearable assistant for Parkinsons disease patients with the freezing of gait symptom. *IEEE Transactions on Information Technology in Biomedicine*, *14*(2), 436–446. https://doi.org/10.1109/TITB.2009.2036165.
Bernstein, N. (1967). The co-ordination and regulation of movement. London: Pergamon.
Bloom, V. et al. (2015). G3Di: A Gaming Interaction Dataset with a Real Time Detection and Evaluation Framework. In *ECCV 2014 Workshops* (pp. 698–712). https://doi.org/10.1007/978-3-319-16199-0
Bloom, V. et al. (2012). G3D: A gaming action dataset and real time action recognition evaluation framework. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 7–12. https://doi.org/10.1109/CVPRW.2012.6239175
Bulling, A. et al. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, *46*(3), 1–33. https://doi.org/10.1145/2499621
Camurri, A. et al. (2016). The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement, 3rd International Symposium on Movement and Computing, MOCO 2016, 5-6 July 2016, Thessaloniki, Greece.
Carletti, V. et al. (2013). Recognition of human actions from RGB-D videos using a reject option. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8158 LNCS*, 436–445. https://doi.org/10.1007/978-3-642-41190-8_47
Carney D.R., et al. (2005). Beliefs about the nonverbal expression of social power. Journal of Nonverbal Behavior, 29:105-123.
Chen, C. et al. (2015). UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. *Proceedings - International Conference on Image Processing, ICIP*, *2015-Decem*, 168–172. https://doi.org/10.1109/ICIP.2015.7350781
Chen, L. et al. (2014). ReadingAct RGB-D action dataset and human action recognition from local features. *Pattern Recognition Letters*, *50*, 159–169. https://doi.org/10.1016/j.patrec.2013.09.004
Cheng, Z. et al. (2012). Human daily action analysis with multi-view and color-depth data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7584 LNCS*(PART 2), 52–61. https://doi.org/10.1007/978-3-642-33868-7_6
de Gelder, B. (2016). Emotions and the Body. New York, Oxford University Press.
De La Torre, F. et al. (2008). Guide to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database. *CMU Robotics Institute*, (April), 19.

Ellis, C. et al. (2013). Exploring the trade-off between accuracy and observational latency in action recognition. *International Journal of Computer Vision*, *101*(3), 420–436. https://doi.org/10.1007/s11263-012-0550-7

Faulkner, H. et al. (2017). TenniSet: A Dataset for Dense Fine-Grained Event Recognition, Localisation and Description. *DICTA 2017 - 2017 International Conference on Digital Image Computing: Techniques and Applications*, *2017-Decem*, 1–8. https://doi.org/10.1109/DICTA.2017.8227494

Fisher, R. B. et al. (2004). The PETS04 surveillance ground-truth data sets. *Evaluation of Tracking and Surveillance*, 1–5. Retrieved from http://homepages.inf.ed.ac.uk/rbf/CAVIAR/PAPERS/pets04.pdf

Fourati, N., & Pelachaud, C. et al. (2014). Emilya : Emotional body expression in daily actions database. *Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 3486–3493.

Gaidon, A. et al. (2012). Mining visual actions from movies To cite this version : Mining visual actions from.

Gkalelis, N. et al. (2009). The i3DPost multi-view and 3D human action/interaction database. *CVMP 2009 - The 6th European Conference for Visual Media Production*, 159–168. https://doi.org/10.1109/CVMP.2009.19

Gorelick, L. et al. (2007). Actions as Space-Time Shapes. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *29*(12), 2247–2253. https://doi.org/10.1109/TPAMI.2007.70711

Goyal, R. et al. (2017). The " something something " video database for learning and evaluating visual common sense. *Proceedings of the IEEE International Conference on Computer Vision*, 5842–5850. https://doi.org/10.1109/ICCV.2017.622

Gross, M. M. et al. (2010). Methodology for Assessing Bodily Expression of Emotion. *Journal of Nonverbal Behavior*, *34*(4), 223–248. https://doi.org/10.1007/s10919-010-0094-x

Gross, R., & Shi, J. et al. (2001). The CMU Motion of Body (MoBo) Dataset. *Cmu-Ri-Tr-01-18*, (June).

Gu, C. et al. (2017). AVA: A Video Dataset of Spatio-temporally Localized Atomic Visual Actions, 6047–6056. https://doi.org/arXiv:1705.08421 [cs.CV]

Hall, J. A. et al (2005). Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. Psychological bulletin, 131 6:898-924

Heilbron, F. C. et al. (2015). ActivityNet: A large-scale video benchmark for human activity understanding. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 961–970. https://doi.org/10.1109/CVPR.2015.7298698

Hicheur, H. et al. (2013). Perception of emotional gaits using avatar animation of real and artificially synthesized gaits. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 460–466. https://doi.org/10.1109/ACII.2013.82

Hochreiter and Schmidhuber, et al. (1997). Long-short-term memory. Neural Comput. 9(8):1735-1780.

Hu, J. F. et al. (2017). Jointly Learning Heterogeneous Features for RGB-D Activity Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(11), 2186–2200. https://doi.org/10.1109/TPAMI.2016.2640292

Hu, T. et al. (2013). Efficient Interaction Recognition through Positive Action Representation. *Mathematical Problems in Engineering*, *2013*, 1–11. https://doi.org/10.1155/2013/795360

Huang, D. et al. (2014). Sequential max-margin event detectors. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8691 LNCS*(PART 3), 410–424. https://doi.org/10.1007/978-3-319-10578-9_27

Huynh, T. et al. (2007). Scalable Recognition of Daily Activities with Wearable Sensors BT - Location- and Context-Awareness. *Location- and Context-Awareness*, *4718*(Chapter 4), 50–67. https://doi.org/10.1007/978-3-540-75160-1_4
Huynh, T. et al. (2008). Discovery of activity patterns using topic models, 10. https://doi.org/10.1145/1409635.1409638
Idrees, H. et al. (2017). The THUMOS challenge on action recognition for videos "in the wild." *Computer Vision and Image Understanding*, *155*, 1–23. https://doi.org/10.1016/j.cviu.2016.10.018
Jacoby N. and McDermott J.H., et al (2017). Intefer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. Current Biology, 27(3):359 - 370.
Jhuang, H.et al. (2013). Towards understanding action recognition. In *Proceedings of the IEEE international conference on computer vision* (pp. 3192–3199).
Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Atten. Percept. Psychophys. 14, 201-211. 10.3758/BF03212378.
Karg, M., & Kirsch, A. et al. (2013). Simultaneous Plan Recognition and Monitoring (SPRAM) for Robot Assistants. In *Human Robot Collaboration Workshop 2013*.
Karpathy, A.et al. (2014). Large-scale video classification with convolutional neural networks. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference On*, 1725–1732. https://doi.org/10.1109/CVPR.2014.223
Kay, W.et al. (2017). The Kinetics Human Action Video Dataset. https://doi.org/10.1145/3123266.3138874
Ke, Y., et al. (2005). Efficient Visual Event Detection Using Volumetric Features. In *Proceedings of International Conference on Computer Vision*. https://doi.org/10.1109/CVPR.2007.383137
Kleinsmith, A. et al. (2011). Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *41*(4), 1027–1038. https://doi.org/10.1109/TSMCB.2010.2103557
Kliper-Gross, O. et al. (2012). The action similarity labeling challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(3), 615–621. https://doi.org/10.1109/TPAMI.2011.209
Kong, Y. et al. (2012). Learning human interaction by interactive phrases. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7572 LNCS*(PART 1), 300–313. https://doi.org/10.1007/978-3-642-33718-5_22
Koppula, H. S. et al. (2013). Learning human activities and object affordances from RGB-D videos. *International Journal of Robotics Research*, *32*(8), 951–970. https://doi.org/10.1177/0278364913478446
Koutník, et al. (2014). A clockwork RNN. In Proceedings of the 31st International Conference on International Conference on Machine Learning Volume 32, CML'14, pages II–1863–II–1871. JMLR.org.
Kuehne, H. et al. (2011). HMDB : A Large Video Database for Human Motion Recognition, 2556–2563.
Kuehne, H. et al. (2014). The Language of Actions : Recovering the Syntax and Semantics of Goal-Directed Human Activities. In *cvpr* (pp. 780–787).
Laptev, I. et al. (2010). Learning realistic human actions from movies.
Laptev, I., & Perez, P. et al. (2007). Retrieving actions in movies.
Li, W.,et al. (2010). Action recognition based on a bag of 3D points. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*, 9–14. https://doi.org/10.1109/CVPRW.2010.5543273

Lillo, I. et al. (n.d.). Discriminative Hierarchical Modeling of Spatio-Temporally Composable Human Activities.
Lin, Y.-C. et al. (2012). Human action recognition and retrieval using sole depth information. In *MM'12* (pp. 1053–1056). https://doi.org/10.1145/2393347.2396381
Liu, A. A. et al. (2015). Coupled hidden conditional random fields for RGB-D human action recognition. *Signal Processing*, *112*, 74–82. https://doi.org/10.1016/j.sigpro.2014.08.038
Liu, A. A. et al. (2015). Multipe/single-view human action recognition via part-induced multitask structural learning. *IEEE Transactions on Cybernetics*, *45*(6), 1194–1208. https://doi.org/10.1109/TCYB.2014.2347057
Liu, P. et al. (2015). Multi-timescale long short-term memory neural network for modelling sentences and documents. Pages 2326-2335.
Ma, Y. et al. (n.d.). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior Research Methods*, *38*(1), 134–141. https://doi.org/10.3758/bf03192758
Mansur, A. et al. (2013). Inverse dynamics for action recognition. *IEEE Transactions on Cybernetics*, *43*(4), 1226–1236. https://doi.org/10.1109/TSMCB.2012.2226879
Meeren, H.K.M et al. (2016). Early preferential responses to fear stimuli in human right dorsal visual stream measured with MEG. Scientific Reports, 6.
Meinard, M. et al. (2007). Mocap Database HDM05.
Messing, R. et al. (2009). Activity recognition using the velocity histories of tracked keypoints. *Proceedings of the IEEE International Conference on Computer Vision*, (Iccv), 104–111. https://doi.org/10.1109/ICCV.2009.5459154
Mettes, P. et al. (2016). Spot on: Action localization from pointly-supervised proposals. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9909 LNCS*(Mil), 437–453. https://doi.org/10.1007/978-3-319-46454-1_27
Mikolajczyk, K., & Uemura, H. et al. (2008). Action recognition with motion-appearance vocabulary forest. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. https://doi.org/10.1109/CVPR.2008.4587628
Montepare, J. M. et al. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior*, *11*(1), 33–42. https://doi.org/10.1007/BF00999605
Munaro, M. et al. (2013). 3D flow estimation for human action recognition from colored point clouds. *Biologically Inspired Cognitive Architectures*, *5*, 42–51. https://doi.org/10.1016/j.bica.2013.05.008
Natarajan, P., & Nevatia, R. et al. (2008). View and scale invariant action recognition using multiview shape-flow models. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. https://doi.org/10.1109/CVPR.2008.4587716
Negin, F. et al. (2013). A decision forest based feature selection framework for action recognition from RGB-depth cameras. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7950 LNCS*, 648–657. https://doi.org/10.1007/978-3-642-39094-4_74
Ni, B. et al. (2011). RGBD-HuDaAct: A color-depth video database for human daily activity recognition. *Proceedings of the IEEE International Conference on Computer Vision*, 1147–1153. https://doi.org/10.1109/ICCVW.2011.6130379

Niebles, J. C. et al. (2010). Modeling temporal structure of decomposable motion segments for activity classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6312 LNCS*(PART 2), 392–405. https://doi.org/10.1007/978-3-642-15552-9_29

Ofli, F. et al. (2013). Berkeley MHAD: A comprehensive Multimodal Human Action Database. *Proceedings of IEEE Workshop on Applications of Computer Vision*, 53–60. https://doi.org/10.1109/WACV.2013.6474999

Oh, S. et al. (2011). A large-scale benchmark dataset for event recognition in surveillance video. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2), 3153–3160. https://doi.org/10.1109/CVPR.2011.5995586

Oreifej, O., & Liu, Z. et al. (2013). HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 716–723. https://doi.org/10.1109/CVPR.2013.98

Patron, A. et al. (2010). High Five: Recognising human interactions in TV shows, 50.1-50.11. https://doi.org/10.5244/c.24.50

Phillips, P. J.et al. (2005). The humanID gait challenge problem: data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(2), 162–177. https://doi.org/10.1109/tpami.2005.39

Pollick, F.E. et al. (2001). Perceiving affect from arm movement. Cognition 82, B51–B61. doi: 10.1016/S0010-0277(01)00147-0.

Ragheb, H. et al. (2008). VIHASI: VIRTUAL HUMAN ACTION SILHOUETTE DATA FOR THE PERFORMANCE EVALUATION OF SILHOUETTE-BASED ACTION RECOGNITION METHODS Hossein Ragheb , Sergio Velastin , Paolo Remagnino and Tim Ellis. In *2008 Second ACM/IEEE Int'l Conference on Distributed Smart Cameras* (pp. 1–10).

Rahmani, H. et al. (2016). Histogram of Oriented Principal Components for Cross-View Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(12), 2430–2443. https://doi.org/10.1109/TPAMI.2016.2533389

Ramanathan, V. et al. (2016). Detecting events and key actors in multi-person videos. https://doi.org/10.1109/CVPR.2016.332

Reiss, A., & Stricker, D. et al. (2012). Introducing a new benchmarked dataset for activity monitoring. *Proceedings - International Symposium on Wearable Computers, ISWC*, 108–109. https://doi.org/10.1109/ISWC.2012.13

Rodriguez, M. D. et al. (2008). Action MACH: A spatio-temporal maximum average correlation height filter for action recognition. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. https://doi.org/10.1109/CVPR.2008.4587727

Roether, C. L. et al. (2009). Critical features for the perception of emotion from gait. *Journal of Vision*, *9*(6), 15–15. https://doi.org/10.1167/9.6.15

Roggen, D. et al. (2010). Collecting complex activity datasets in highly rich networked sensor environments. *INSS 2010 - 7th International Conference on Networked Sensing Systems*, (00), 233–240. https://doi.org/10.1109/INSS.2010.5573462

Rohrbach, A. et al. (2015). A dataset for Movie Description. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 3202–3212. https://doi.org/10.1109/CVPR.2015.7298940

Rohrbach, M. et al. (2012). A database for fine grained activity detection of cooking activities. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1194–1201. https://doi.org/10.1109/CVPR.2012.6247801

Rohrbach, M. et al. (2016). Recognizing Fine-Grained and Composite Activities Using Hand-Centric Features and Script Data. *International Journal of Computer Vision*, *119*(3), 346–373. https://doi.org/10.1007/s11263-015-0851-8

Ryoo, M. S., & Aggarwal, J. K. et al. (2009). Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. *Proceedings of the IEEE International Conference on Computer Vision*, (Iccv), 1593–1600. https://doi.org/10.1109/ICCV.2009.5459361

Sapp, B. et al. (2011). Parsing human motion with stretchable models. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1281–1288. https://doi.org/10.1109/CVPR.2011.5995607

Schüldt, C. et al. (2004). Recognizing human actions : A local SVM approach Recognizing Human Actions. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.* (Vol. 3, pp. 3–7). https://doi.org/10.1109/ICPR.2004.1334462

Seidenari, L. et al. (2013). Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 479–485. https://doi.org/10.1109/CVPRW.2013.77

Shahroudy, A. et al. (2016). NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis, 1010–1019. https://doi.org/10.1109/CVPR.2016.115

Shimada, A. et al. (2012). Kitchen Scene Context Based Gesture Recognition: A Contest in ICPR2012 see p. 177. In X. Jiang, O. Regina, P. Bellon, D. Goldgof, & T. Oishi (Eds.), *Advances in Depth Image Analysis and Applications*. Springer. https://doi.org/10.1007/978-3-642-40303-3

Shoaib, M. et al. (2014). *Fusion of smartphone motion sensors for physical activity recognition*. *Sensors (Switzerland)* (Vol. 14). https://doi.org/10.3390/s140610146

Sigal, L. et al. (2010). HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International Journal of Computer Vision*, *87*(1–2), 4–27. https://doi.org/10.1007/s11263-009-0273-6

Sigurdsson, G. A. et al. (2016). Hollywood in homes: Crowdsourcing data collection for activity understanding. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9905 LNCS*, 510–526. https://doi.org/10.1007/978-3-319-46448-0_31

Singh, B. et al. (2016). A Multi-stream Bi-directional Recurrent Neural Network for Fine-Grained Action Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1961–1970. https://doi.org/10.1109/CVPR.2016.216

Singh, S. et al. (2010). MuHAVi: A multicamera human action video dataset for the evaluation of action recognition methods. *Proceedings - IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010*, 48–55. https://doi.org/10.1109/AVSS.2010.63

Slowinski et al. (2016). Dynamic similarity promotes interpersonal coordination in joint-action. Journal of the Royal Society Interface, 13: 20151093. http://dx.doi.org/10.1098/rsif.2015.1093.

Song, Y. et al. (2014). Body surface context: A new robust feature for action recognition from depth videos. *IEEE Transactions on Circuits and Systems for Video Technology*, *24*(6), 952–965. https://doi.org/10.1109/TCSVT.2014.2302558

Soomro, K. et al. (2012). UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild, (November). https://doi.org/10.1016/S0021-9290(07)70002-7
Stiefmeier, T. et al. (2008). Wearable Activity Tracking in Car Manufacturing. *IEEE Pervasive Computing*, *7*(2), 42–50. https://doi.org/10.1109/mprv.2008.40
Stisen, A.et al. (2015). Smart Devices are Different, (December 2016), 127–140. https://doi.org/10.1145/2809695.2809718
Sun, C. et al. (2015). Temporal Localization of Fine-Grained Actions in Videos by Domain Transfer from Web Images, 371–380. https://doi.org/10.1145/2733373.2806226
Sung, J. et al. (2011). Human Activity Detection from RGBD Images. *Artificial Intelligence*, *abs/1107.0*, 47–55. https://doi.org/citeulike-article-id:9510173
Sung, J. et al. (2012). Unstructured human activity detection from RGBD images. *Proceedings - IEEE International Conference on Robotics and Automation*, 842–849. https://doi.org/10.1109/ICRA.2012.6224591
Tang, C. et al. (2018). Online human action recognition based on incremental learning of weighted covariance descriptors. *Information Sciences*, *467*, 219–237. https://doi.org/10.1016/j.ins.2018.08.003
Tenorth, M. et al. (2009). The TUM Kitchen Data Set of Everyday Manipulation Activities for Motion Tracking and Action Recognition Moritz. In *2009 IEEE 12th International Conference on Computer Vision Workshops* (pp. 1089–1096). Retrieved from http://www9.cs.tum.edu/publications/pdf/tenorth09dataset.pdf
Theodorakopoulos, I. et al. (2014). Pose-based human action recognition via sparse representation in dissimilarity space. *Journal of Visual Communication and Image Representation*, *25*(1), 12–23.
Tilmanne, J. et al. (2008). A Database for Stylistic Human Gait Modeling and Synthesis see p. 91. In *ENTERFACE '08*.
Tilmanne, J., & Dutoit, T. et al. (2010). Expressive Gait Synthesis Using PCA and Gaussian Modeling. In R. Boulic, Y. Chrysantou, & T. Komura (Eds.), *Lecture Notes in Computer Science* (Vol. 6459, pp. 363–374). Berlin Heidelberg: Springer-Verlag.
Torabi, A. et al. (2015). Using Descriptive Video Services to Create a Large Data Source for Video Annotation Research. Retrieved from http://arxiv.org/abs/1503.01070
Um, T. T. et al. (2017). Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring using Convolutional Neural Networks. https://doi.org/10.1145/3136755.3136817
van Gemeren, C. et al. (2014). Dyadic Interaction Detection from Pose and Flow, 101–115. https://doi.org/10.1007/978-3-319-11839-0_9
Vaswani, et al. (2017). Attention is all you need. In NIPS.
Veeraraghavan, A. et al. (2006). The function space of an activity. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *1*, 959–966. https://doi.org/10.1109/CVPR.2006.304
Wallbott, H.G. et al. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, *28*(6), 879--896.
Wang, J. et al. (2012). Mining actionlet ensemble for action recognition with depth cameras. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1290–1297. https://doi.org/10.1109/CVPR.2012.6247813

Wang, J. et al. (2014). Cross-view action modeling, learning, and recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2649–2656. https://doi.org/10.1109/CVPR.2014.339

Wang, K. et al. (2015). 3D Human Activity Recognition with Reconfigurable Convolutional Neural Networks. https://doi.org/10.1145/2647868.2654912

Wang, P. et al. (2016). Action Recognition from Depth Maps Using Deep Convolutional Neural Networks. *IEEE Transactions on Human-Machine Systems*, *46*(4), 498–509. https://doi.org/10.1109/THMS.2015.2504550

Wang, P., & Rehg, J. M. et al. (2006). A modular approach to the analysis and evaluation of particle filters for figure tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *1*, 790–797. https://doi.org/10.1109/CVPR.2006.32

Wang, X. et al. (2016). Actions ~ Transformations. In *Cvpr* (pp. 2658–2667).

Wei, P. et al. (2013). Modeling 4D human-object interactions for event and object recognition. *Proceedings of the IEEE International Conference on Computer Vision*, 3272–3279. https://doi.org/10.1109/ICCV.2013.406

Wei, P. et al. (2013). Concurrent action detection with structural prediction. *Proceedings of the IEEE International Conference on Computer Vision*, (1), 3136–3143. https://doi.org/10.1109/ICCV.2013.389

Weinland, D. et al. (2006). Free viewpoint action recognition using motion history volumes. *Computer Vision and Image Understanding*, *104*(2-3 SPEC. ISS.), 249–257. https://doi.org/10.1016/j.cviu.2006.07.013

Weinzaepfel, P. et al. (2017). Towards Weakly-Supervised Action Localization.

Wolf, C. et al. (2014). Evaluation of video activity localizations integrating quality and quantity measurements. *Computer Vision and Image Understanding*, *127*, 14–30. https://doi.org/10.1016/j.cviu.2014.06.014

Wu, C. et al. (2015). Watch-n-Patch : Unsupervised Understanding of Actions and Relations. In *Cvpr* (pp. 4362–4370).

Xia, L. et al. (2012). View invariant human action recognition using histograms of 3D joints. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 20–27. https://doi.org/10.1109/CVPRW.2012.6239233

Xu, J. et al. (2016). MSR-VTT: A Large Video Description Dataset for Bridging Video and Language, 5288–5296. https://doi.org/10.1109/cvpr.2016.571

Xu, N. et al. (2015). A Multi-modal & Multi-view & Interactive Benchmark Dataset for Human Action Recognition Categories and Subject Descriptors. In *MM'15* (pp. 1195–1198).

Yeffet, L., & Wolf, L. et al. (2009). Local trinary patterns for human action recognition. *Proceedings of the IEEE International Conference on Computer Vision*, (Iccv), 492–497. https://doi.org/10.1109/ICCV.2009.5459201

Yeung, S. et al. (2017). Every Moment Counts: Dense Detailed Labeling of Actions in Complex Videos. *International Journal of Computer Vision*, *126*(2–4), 375–389. https://doi.org/10.1007/s11263-017-1013-y

Yun, K. et al. (2012). Two-person interaction detection using body-pose features and multiple instance learning. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 28–35. https://doi.org/10.1109/CVPRW.2012.6239234

Zhang, C., & Tian, Y. et al. (2012). RGB-D Camera-based Daily Living Activity Recognition. *Journal of Computer Vision and Image Processing*, *2*, 12. Retrieved from http://media-lab.ccny.cuny.edu/wordpress/Publications/NWPJ-201209-15-CameraReady.pdf

Zhang, J. et al. (2018). A large scale RGB-D dataset for action recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10188 LNCS*(December), 101–114. https://doi.org/10.1007/978-3-319-91863-1_8

Zhang, W. et al. (2013). From actemes to action: A strongly-supervised representation for detailed action understanding. *Proceedings of the IEEE International Conference on Computer Vision*, 2248–2255. https://doi.org/10.1109/ICCV.2013.280

Zhang, Z. et al. (2012). A Viewpoint-Independent Statistical Method for Fall Detection. *International Conference on Pattern Recognition*, (Icpr), 3626–3630.

Zhao, H. et al. (2017). HACS: Human Action Clips and Segments Dataset for Recognition and Temporal Localization. Retrieved from http://arxiv.org/abs/1712.09374