

D1.3 Predictions of multi-scale, temporally aggregated, social-motor synchronization models - 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. SENSORIMOTOR COMMUNICATION	3
1.1 THE NEURO-FUNCTIONAL SUBSTRATE OF SENSORIMOTOR COMMUNICATION	4
1.2 THE IDIOSYNCRATIC NATURE OF ACTION EXECUTION AND ACTION PERCEPTION MATCHING	5
1.3 THE NEUROPHYSIOLOGY OF ACTION PERCEPTION DURING INTERACTION	7
2. THE USE OF MUSICIANS TO STUDY SENSORIMOTOR PROCESSES	7
2.1 THE USE OF ENSEMBLE MUSICIANS TO STUDY SENSORIMOTOR COMMUNICATION	8
2.2 THE CASE OF ORCHESTRAS AND QUARTETS	11
2.3 SENSORIMOTOR COMMUNICATION	15
3. MOVEMENT MARKERS OF EMOTIONAL QUALITIES	15
3.1 INTRODUCTION TO IMS AND IEMS	17
3.2 MULTI-TIME MODE CONCEPT	18
4. MACHINE LEARNING FOR MOVEMENT ANALYSIS	19
4.1 DISCOVERING RELEVANT MOVEMENT PATTERNS	20
4.2 PREDICTING MULTIPLE FUTURES	21
4.3 ABSTRACTING REPRESENTATIONS OF MOVEMENTS	22
4.4 COUPLING AND TRANSFER OF INFORMATION	23
4.5 PROVIDING INSIGHT INTO THE DATA AVAILABLE AND THE NEED FOR MORE DATA	23
5. INTRODUCTION TO EXPERIMENTAL WORK	25
5.1 INDIVIDUAL LEVEL	25
5.2 DYADIC LEVEL	26
5.3 GROUP LEVEL	27
6. REFERENCES	29

1. SENSORIMOTOR COMMUNICATION

All animal species grouping for defensive, reproductive or hunting needs have evolved complex communicative behaviors to achieve coordinated action (Frith 2008)(Rands et al. 2003)(Couzin et al. 2005)(Nagy et al. 2010). Humans are innately social creatures and there is little doubt that cognition and brain organization are shaped around this fact. Nevertheless, cognitive neuroscience took quite some time to acknowledge that the study of cognition in isolated individuals might be an ill-posed scientific approach. Recently however, a clear switch towards what has been called “second person neuroscience” has led to an apparent blossoming of the study of the brain during real-time truly interactive scenarios (Schilbach et al. 2013).

A major theoretical shift in this regard was driven by a new conceptualization of the motor system. The motor system was once believed to be an output system, slavishly following the dictate of the perceptual brain. However, motor processes seem to also play a role in perceptual and cognitive functions, challenging the classical sensory versus motor separation and opening the doors to embodied cognition research in both humans and artificial systems (Clark and Grush 1999). In addition to the fundamental role in movement planning and execution, some premotor neurons show also complex visual responses. Among them, “mirror neurons” discharge both when the monkey executes an action and observes another individual make the same action (Gallese et al. 1996). On the other hand, “canonical neurons” are activated both during object-directed actions and object observation alone (Murata et al. 1997). These discoveries have stimulated a large wealth of basic monkey and human research and after about 20 years we know a great deal of details regarding the role and function of the motor system in action and object perception (D’ausilio, Bartoli, and Maffongelli 2015a).

In fact, action properties and object geometry carry reliable statistical features we can exploit for classification. Indeed, actions must comply with the laws of biological motion and body biomechanics, whereas the shape of objects/tools constrain the interaction potentialities allowed to the human hands. All this a priori knowledge offers a great advantage for any system that tries to optimize behavioral interaction. In this sense, activities in the motor system are not intended as a coding schema of the external world but rather as a generative engine to extract different levels of features depending on context (D’ausilio, Bartoli, and Maffongelli 2015b). Critically, the recruitment of motor programs, during action/object perception, constrain the active search of specific sensory features that maximize the discrimination between two perceptual hypotheses and/or support prediction of future information (Friston, Mattout, and Kilner 2011).

The generation of active inferences about future actions of conspecifics is central to our capability to smoothly interact with each other and, therefore, fundamental to the development of human cognition. An interesting research theme revolves around the investigation of how these mechanisms are instantiated at the brain level and how these processes are used to facilitate human-to-human interaction. In simple terms, the quest for the understanding of how, why and when we do send and receive implicit sensorimotor messages during complex social interaction.

The mirror fronto-parietal circuit, due to its properties, might be particularly important in communication between individuals, especially when coordinated action is central (Hurley 2008). In fact, the other’s action might be first represented in our motor system via mirror-like mechanisms, such that we can easily have access to “how” the agent is doing what he’s doing (Fadiga et al. 1995). This knowledge, on one hand, will help us understand the agent’s intentions via additional inferential processes eventually based on experience and contextual information (Marco Iacoboni et al. 2005). On the other hand, instead, this low level sensory-motor resonance with our own motor planning repertoire

might enable efficient prediction, anticipation and planning of appropriate actions in response or modify our behavior online (M Iacoboni et al. 1999).

Action mirroring, however, does not facilitate coordinated action per se, in fact, coordination may often require the execution of very different actions between participants. For example, if we move a heavy load together with somebody else we might be better tracking the other's action in a fast and efficient manner. In fact, if our fellow produces a particular movement in one direction we need to compensate with a concurrent and opposite action that takes into account several kinematic and dynamic aspects of the situation. This is a clear example of coordinated action where one individual is temporarily causing the others' behavior (Sebanz, Bekkering, and Knoblich 2006). Recently however, it has been shown that complementary action observation may recruit human mirror neuron areas to a greater extent (Newman-Norlund et al. 2007) thus suggesting that the human mirror system might be tuned to action coordination rather than rote action mirroring. We propose that the mirror-mechanism might provide the means to represent the other's action in motor coordinates rather than visual or symbolic (i.e. language) that in turn may result extremely useful in anticipating other's actions, planning a motor correction and hence be useful for fast coordinated behaviors. In this context, we might use our motor internal models to gain access to low-level control parameters implemented by the people interacting with us. Therefore, action coordination may benefit from the ability to model others' behavior implementation and use it as an additional prior, in a Bayesian perspective (Friston, Mattout, and Kilner 2011).

1.1 The neuro-functional substrate of sensorimotor communication

This sensory-motor conversation requires that all participants must be able to send and receive subtle messages in the form of visual motor gestures and auditory events. Therefore, critical for this account is the ability to encode/decode such messages. This encoding/decoding process might be conceived as a complex and hierarchical input-output mapping ranging from rote sensory-motor mapping to the highest level of human action organization (Grafton and Hamilton 2007). The fronto-parietal networks are the best neural candidates for this computation since they are characterized by extensive structural connections (Petrides and Pandya 2009) supporting critical higher-level sensori-motor coordinate transformations (Giacomo Rizzolatti and Sinigaglia 2016). The circuit connecting the anterior intraparietal sulcus (AIP) and the posterior bank of monkey inferior arcuate sulcus (F5), for instance, has shown interesting functional properties (Borra et al. 2017). A subset of neurons in both these areas was activated by the execution of goal-directed actions and the observation of kinematically similar actions performed by other individuals (Gallese et al. 1996)(Fogassi et al. 2005).

Originally found in monkeys by using single unit recordings, subsequent studies have shown similar mechanisms in humans as well. Different techniques such as transcranial magnetic stimulation (Fadiga et al. 1995), fMRI or PET (Rizzolatti et al. 1996) and MEG (Hari et al. 1998) have confirmed these results. More recently, a large body of research is also describing these mechanisms in humans by using novel methods and approaches (Gazzola and Keysers 2009)(Kilner et al. 2009).

Immediately following the initial reports of mirror neurons in the macaque brain, the existence of an analogous mechanism in humans was confirmed by numerous results coming from various techniques such as transcranial magnetic stimulation (TMS; (Naish et al. 2014))electroencephalography (EEG;(Fox et al. 2016)), functional magnetic resonance imaging (fMRI; (Hardwick et al. 2018)) and human single-cell recordings ((Mukamel et al. 2010)) revealed the existence of a fronto-parietal network with mirror-like properties in humans ((Giacomo Rizzolatti and Sinigaglia 2016)).

Based on human brain-imaging data (Hardwick et al. 2018) and cytoarchitecture (Petrides, Cadoret, and Mackey 2005), the ventral premotor cortex and the pars opercularis of the posterior inferior frontal gyrus (Brodmann area 44) were assumed to be the human homologues of macaque mirror area F5. Later, the rostral inferior parietal lobule was identified as equivalent of the monkey mirror area PF/PFG (Giacomo Rizzolatti and Sinigaglia 2016).

In parallel, EEG research showed that event-related synchronization and desynchronization of the *mu* rhythm (Rolandic alpha band) were linked to action performance, observation and imagery (Fox et al. 2016). These results suggest that Rolandic *mu* event-related desynchronization during action observation reflects activity of a mirror-like system present in humans (Fox et al. 2016).

Finally, single-pulse TMS over the primary motor cortex (M1) and motor evoked potentials (MEPs) amplitude were employed as a direct index of corticospinal recruitment (Corticospinal Excitability - CSE). Using this technique, several studies showed a modulation of MEPs amplitude during action observation matching various changes occurring during action execution (Naish et al. 2014).

Although all these measures have been instrumental to investigate mechanisms underlying action observation in humans, the most useful to determine the degree of matching between observed and executed action remains TMS. Differently from other techniques that either measure slow metabolic signals (e.g. PET, SPECT, fMRI, fNIRS) or the electrical mass activity of an extended brain networks (EEG-MEG), TMS allows the direct measurement of motor activity. In fact, a single pulse of TMS noninvasively stimulate the human motor cortex to instantaneously assesses the magnitude of the descending motor drive to muscles (Hallett 2007). MEP size reflects the net facilitatory and inhibitory input to the pyramidal projecting neurons (Rothwell 1997), thus providing an instantaneous read-out of the activity of an extended motor network.

This approach has been classically used to evaluate corticospinal excitability during motor imagery (Fadiga et al. 1999), action observation (Naish et al. 2014) as well as planning or execution of an action (Hoshiyama et al. 1996). The high temporal resolution of the technique allows effective exploration of motor recruitment in these tasks. Indeed, this technique provides direct comparison between the unfolding of motor processes during action observation and the timing of real muscle activation during action execution (Prabhu et al. 2007). Until today, more than a hundred TMS studies have shown that the neural match between action execution and observation (Amoruso and Finisguerra 2019) is characterized by an important degree of temporal and somatotopic congruency between the motor representations elicited by two conditions. Indeed, MEPs are modulated by observed low-level movement features such as finger aperture in a grasping action (Gangitano, Mottaghy, and Pascual-Leone 2001), the amplitude of muscle activities over time (Borroni et al. 2005) and the forces required to lift objects of different weights (Alaerts et al. 2010). In parallel, modulation of CSE amplitude was also shown for higher level features such as action goals (Cattaneo et al. 2009). For example, MEPs amplitude did not seem to depend on the effector used to attain the same action goal (Borroni et al. 2005)(Finisguerra et al. 2015), suggesting independence from low-level movement features.

1.2 The idiosyncratic nature of action execution and action perception matching

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 (Flanagan and Johansson 2003)(Falck-Ytter, Gredeback, and von Hofsten 2006) or infer higher-order cognitive processes hiding behind those actions (Becchio et al. 2008)(Soriano et al. 2018). In principle, knowledge about the invariant

properties of movement control (Flash and Hogan 1985) could support inferences about the unfolding of other's actions (Casile et al. 2010)(Dayan et al. 2007). In this regard, it has been proposed that these inferences may be based on a direct match between actor's sensorimotor activations during Action Execution (AE) and observer's sensorimotor activations triggered by Action Observation (AO). Indeed, using Corticospinal Excitability (CSE), motor recruitment during AO was shown to replicate the spatio-temporal sequence of motor commands implemented by the actor (for a review please see: (Naish et al. 2014).

This idea is however challenged by the redundancy that characterizes the organization of human movement (Pauline M Hilt et al. 2017). 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 (Bernstein 1967). In this regard, a strong version of the direct-matching hypothesis (Giacomo Rizzolatti and Sinigaglia 2016) explains inferences when a direct relationship exists between muscle recruitment, movement kinematics and behavioral goals (e.g. simple finger movements). However, it is less clear how other's complex movements (i.e. multi-joint movements) are transformed onto the observer's motor representations. In this case, any sensorimotor-based inference about other's actions, amount to finding a solution to a many-to-many mapping problem.

Recently it was shown that a simpler mapping exists between behavioral goals and the lower dimensionality space of whole-body configurations (i.e. synergies;(Pauline M Hilt et al. 2017)). In fact, although a handful of kinematic solutions are biomechanically valid, everyday actions (i.e. reaching for an object on the floor starting from a standing posture) are usually performed via a limited number of possible kinematic configurations of the biomechanical chain (e.g. "ankle" and "hip" strategies for postural control; (Horak and Nashner 1986)). On the top of that, each individual carry his own robust and yet unique way of moving (Individual Motor Signature – IMS; (Hilt et al. 2016)(Slowinski et al. 2016). For instance, in a whole-body reaching task Hilt and collaborators (Hilt et al. 2016) showed low intra-subject motor variability, accompanied by a large inter-subject variability. The inherent lower dimensionality of whole-body postural control and the presence of robust Individual Motor Strategies (IMS) suggest the existence of a simpler AO-AE mapping that may be function of everyone's individual movement style.

Individual differences in the execution of a multi-joint action shape the sensorimotor activities during the observation of the same action (Hilt et al., In Press). This shaping is made visible by our experimental design but should in principle be an ingredient of any multi-joint action. Beside the general suggestion that inter-subject variability should be considered as a tool rather than a problem, our results force us to redefine the core properties of the motor simulative account. At the same time, we acknowledge that showing generalization to different motor domains or action types is of critical importance. In this regard, a strong parallel can already be drawn with findings in the speech domain. Here, the degree of motor recruitment during listening to syllables scales for the perceived distance between listener and speaker (Bartoli et al. 2015). Moreover, lip corticobulbar excitability during speech listening is greater for speech sounds that are far from the listener's motor repertoire (Schmitz et al. 2019). Supported by converging results from two relatively far domains of action-perception integration, it was proposed that the AO effects reflects sensitivity to differences rather than similarities with respect to other's behavior.

1.3 The neurophysiology of action perception during interaction

Sensorimotor activity during AO may support action-related perceptual processes (Avenanti, Candidi, and Urgesi 2013). 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 (Donnarumma et al. 2017)(Friston, Mattout, and Kilner 2011). Perceptual discrimination and prediction of other's actions, may have a key role in supporting temporal and spatial interpersonal coordination (Pezzulo et al. 2019). 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 (Silvanto, Muggleton, and Walsh 2008). We can thus expect that temporal and spatial overlap of the neural processes subtending AE and AO produces functionally relevant interaction.

In fact, behavioral interaction in natural settings occurs at fast pace and humans coordinate their actions by quickly adapting to other's behavior. This means that neural processes subtending AE and AO can unfold smoothly, notwithstanding their important temporo-spatial overlap (Novembre et al. 2014). However, at the behavioral level AE interferes with the process of visual action recognition (de la Rosa, Ferstl, and Bulthoff 2016). Proactive eye movements, which are present during visually guided actions and during AO (Flanagan and Johansson 2003), are reduced when an AO-AE mismatch is present (Costantini, Ambrosini, and Sinigaglia 2012). Similarly, the observation of objects affording a specific grasp, biases concurrent grasping performances (Rounis, van Polanen, and Davare 2018). In general, AE is facilitated by compatible and impeded by incompatible AO (Cracco et al. 2018). These results suggest that the neural processes subtending AO and AE modulate each other. A recent neurophysiological study ((Cardellicchio et al. 2020)) considered participants as actors and observers at the same time, in fact they produced a tonic motor descending drive, while observing others' actions. Corticospinal inhibition decreased during mismatching executed and observed actions. These results demonstrate that AE-AO mismatch is computed at the level of muscle recruitment and according to an agonist-antagonist mapping of actions. Critically, the functional contribution of muscles to a specific action seems to be the guiding principle in allowing modulation of corticospinal inhibitory circuits for AE-AO mismatching conditions.

2. THE USE OF MUSICIANS TO STUDY SENSORIMOTOR PROCESSES

While engaged in naturalistic joint action, action coordination may be subject to data availability (full body vision Vs occluded vision), context richness (known vs novel context) or expertise in a given task (Giacomo Rizzolatti and Sinigaglia 2016). If on one hand it is clear that partly occluded vision or the presence of a reliable context reduces the impact of low-level decoding of others' action kinematics, the case of experts is particularly interesting. In fact, sport, dance or music experts do not obtain a simple rough interpretation of others' actions. Rather, expertise in these activities is all about anticipatorily modeling of low-level movement features (Aglioti et al. 2008).

In fact, professional musicians, for instance, undergo important plastic changes induced by extensive motor and sensory training (Elbert et al. 1995)(Schlaug et al. 1995)(Pantev et al. 1998). These anatomo-functional changes are also paralleled by enhanced ability to discriminate subtle changes in others' performance via predictive action simulation (A D'Ausilio et al. 2006)(Candidi et al. 2014).

One facet of real expertise may be the refinement of sensori-motor skills as well as the domain-specific tuning of low-level sensori-motor feature extraction from others behavior. In this vein, musicians have been extensively used with these purposes. For example, drummers were used to study time processing (Cicchini et al. 2012), violinist players for somatosensory plasticity (Elbert et al. 1995), piano players for sensori-motor integration (A D'Ausilio et al. 2006) or jazz players to study creativity and improvisation (Limb and Braun 2008).

For instance, listening to music evokes specific sensorimotor activity in musicians. This fact has been shown using various neuroimaging and neurophysiological techniques (MEG: (Hauelsen and Knosche 2001); EEG: (Bangert and Altenmuller 2003); fMRI: (Lotze et al. 2003); TMS: (A D'Ausilio et al. 2006)). These results broadly suggest that a fronto-parietal network of brain areas, usually activated by action planning and execution, is, at least partially, recruited in the musicians' brain when listening to rehearsed musical segments. Similarly, the visual presentation of musical actions (e.g. videos of hand playing a piano) evokes specific motor brain responses (Candidi et al. 2014), suggesting that musical training also generates new visuo-motor associations. Importantly, these sensorimotor activities are triggered by sensory stimuli, even when no motor response is needed (Zatorre, Chen, and Penhune 2007), suggesting that long-term training produces stable sensorimotor associations. Therefore, the motor system in musicians show mirror-like properties and it seems plausible that this mechanism could support their orchestrated musical execution.

2.1 The use of ensemble musicians to study sensorimotor communication

Recently it has been proposed that music can offer a unique solution to balance experimental control and ecological testing of cognition and brain functions (D'Ausilio et al. 2015). More interestingly, musicians especially ensemble instrumentalists, are experts in a form of social interaction characterized by real-time non-verbal communication. Ensemble musicians train for years in order to refine skills that allow them to accurately encode and decode subtle sensorimotor non-verbal messages with the main purpose of establishing and maintaining a shared coordinative goal. In group level musical coordination individuals might be conceptualized as processing units embedded within a complex system (i.e. the ensemble) and sharing an aesthetic and emotional goal. Each participant may thus non-verbally transmit sensory information while in parallel decoding other's movement. The sensory information generated by the sender is based on body movements and is transferred through the visual (i.e., body sway, head motion), auditory (i.e., instrument sounds) and somatosensory channels (i.e., floor vibrations). With information flowing, participants may rely on predictive models to cope with the real-time demands of interpersonal coordination. Thus, the musical ensemble constitutes a dynamical system that possesses important constraints due to the fact that co-performers behave in a complex but formalized manner dictated by musical conventions and often a musical score.

As a matter of fact, for musical ensemble playing the complexity of behavioural coordination can be formalized as the accurate control of relative movement timing during performance. For example, quartets exhibit a near-optimal gain, whereas individual gains reflected contrasting strategies of first-violin-led autocracy versus democracy group organization (Wing et al. 2014). Furthermore, the familiarity with a co-performer's part has dissociated effects on different levels of movement control. Keystroke coordination was affected by predictions about expressive micro timing, based on the performer's own playing style rather than the co-performer's style. Body sway coordination was, instead, modulated by temporal predictions related to musical phrase structure (Ragert, Schroeder, and Keller 2013). These studies are starting to shed some initial light on the possibility that complex

behavioral coordination could be based on a hierarchy of body motion control, each differentially modulated by social and/or situational factors. In this sense, ensemble music seems to be an effective test-bed for such a domain-general capability.

A recent project, Interpersonal Entrainment in Music Performance (IEMP), proposes a distinction between entrainment processes at different time scales: at shorter time-scales (100s of milliseconds), sensorimotor synchronization processes ensure appropriate alignment of musical parts, while at longer time-scales (several seconds) musicians use different processes to align at key points in time and manage transitions (e.g. between different textures or tempi). In this model, although synchronization requires musicians to share an understanding of the piece or at least relevant stylistic parameters, its accuracy depends largely on physical factors such as the rate of event onsets and their envelope characteristics (thus in a comparative study, fast music and that for drums/percussion is measured to be synchronized more accurately than slower music and/or that by instruments with less well-defined attacks; Clayton et al. submitted). Longer-term coordination processes are more dependent on conscious management of performance by ensemble musicians, can involve explicit cueing, and can often be detected via greater coherence in body sway as measured using Cross-Wavelet Transform (CWT) analysis of movement extracted from performance videos (Eerola et al. 2018). A comparative study of these two processes as exemplified in a corpus of Indian instrumental music performances highlights their complementarity (Clayton et al. 2019). In this study CWT analysis showed that movement coherence between musician duos is significantly greater on metrical downbeats, particularly cadential downbeats that mark the end of improvisatory episodes, while metrical structure has no consistent effect on synchronization accuracy.

On the other hand, a series of studies based on motion kinematic data acquired from quartets or orchestras, used a complementary computational approach. These studies, recorded musicians playing in an ecological scenario and employed mathematical tools to extract natural information flow between participants. These data-driven techniques, such as the Granger Causality method, showed that the increase of conductor-to-musicians influence, together with the reduction of musician-to-musician coordination was associated with quality of execution, as assessed by musical experts ((A D'Ausilio et al. 2012); Figure 1). A similar methodology was also able to extract implicit dominance and leadership in quartets (Badino et al. 2014), whereas entropy-based measures distinguished between solo and ensemble performances (Glowinski et al. 2013). Together, these studies suggest that using specific computational tools it is possible to blindly extract the pattern of information transfer between participants engaged in a realistic and complex interaction – without any variables manipulation and behavioural interference.

According to the mirror-matching hypothesis, one possible mechanism supporting behavioural coordination via sensorimotor information transfer is the motor simulation of other's behaviour. For instance, the role of motor simulation in inter-personal coordination was investigated in pianists listening to a recording of a piece trained with the left hand, and asked to play a complementary part with the right hand. Interference with the motor cortex, impaired inter-personal coordination only when the heard pieces had been trained (Novembre et al. 2012). Of further interest is the finding that the mere presence of a co-performer modulated cortico-spinal excitability in amateur pianists asked to perform the right-hand part of a musical piece, while the complementary left-hand part was believed to be performed by a hidden co-performer (Novembre et al. 2012). This kind of studies might constitute a powerful tool to study the neurophysiological role of other's action simulation during dynamic interaction and behavioural coordination.

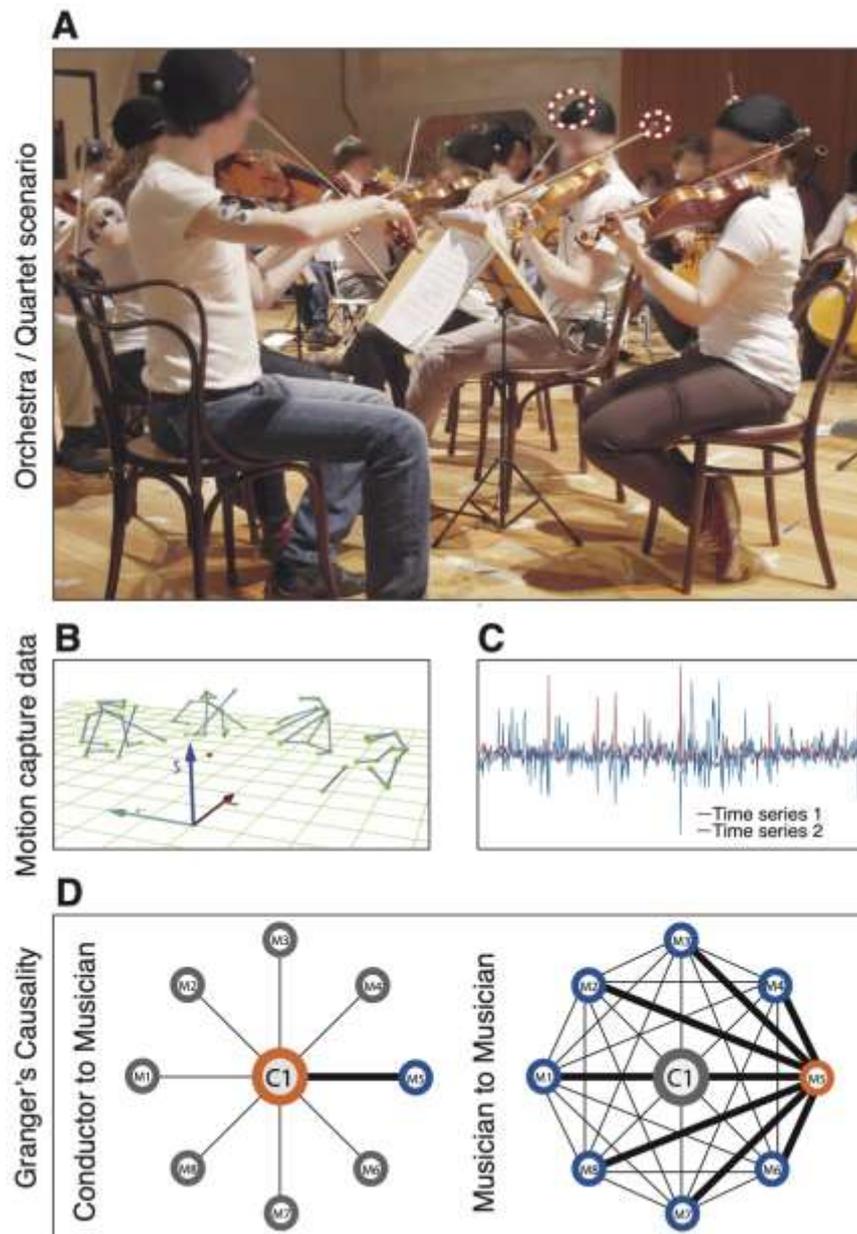


Figure 1. Panel A shows a picture of a motion capture recording set-up. Highlighted the markers on the musician's head and bow's. In panel B, a stick-figure representation of a quartet performing a musical piece. In panel C, two time-series representing accelerometric data from a motion capture recording. In panel C, a graphical description of the metrics that can be extracted from movements. Specifically, on the left side what we called "Conductor to Musicians" describe the causal drive from the conductors towards all musicians. On the right side, what we named "Musicians to Musicians", an index that while removing the statistical contribution of the conductor measured the inter-musician sensorimotor communication.

At the same time other recent studies started the investigation of musical interaction by complementing behavioural measures with classical metabolic and electrical neuroimaging. For instance, it was shown that simultaneous electroencephalographic recording in musicians playing in ensemble could be achieved with a limited amount of motion artefacts (Babiloni et al. 2011). Following this approach it

was possible to dissociate the processes related to monitoring the self's performance and the joint action outcome (Loehr et al. 2013). Even more compelling is the possibility to investigate the amount of inter-brain information flow. In fact, patterns of directed between-brain coupling, in alpha and beta frequency ranges, was associated with the musical roles of leader and follower (Sanger, Muller, and Lindenberger 2013). Also, the increase of inter-brain phase coherence in delta and theta frequency bands was increased in frontal and central scalp regions during periods that presented higher demands on coordination (Sanger, Muller, and Lindenberger 2012).

Functional Magnetic Resonance Imaging showed the neural correlates of inter-personal synchrony and its effect on pro-social behaviour. Synchronization was shown to increase activity in the caudate (reward system) and predicted prosodic behaviour (Kokal et al. 2011). Another study identified distinct cortical networks that were selectively activated depending on the kind of cooperativeness of an interaction partner. Cooperative virtual partners who corrected for moderate amounts of synchronization error facilitated performance and were associated with the activation of cortical midline structures linked to socio-affective processes. Virtual partners who over-compensated for errors led to poor inter-agent synchronization and the activation of lateral prefrontal areas associated with executive functions and cognitive control (Uhlir, Fairhurst, and Keller 2013). Also, leader-follower dynamics in inter-agent coordination showed that agency related brain regions were strongly activated in leaders, consistent with their greater self-focus (Fairhurst, Janata, and Keller 2014).

In summary, the basic building blocks enabling social dominance, leadership, or cooperativeness might be effectively be studied in ensemble musicians through the use of multi-dimensional approaches (involving combinations of brain imaging and stimulation, kinematic measures of large-scale body movements, and measures of sensorimotor synchronization at the millisecond time scale). These basic social and cognitive processes can be generalized to similar, not necessarily musical, cognitive phenomena. In fact, the human capacity for music might have evolved as a tool that fosters social bonding and group cohesion in general (Kirschner and Tomasello 2010).

Group level musical coordination can be considered as a microcosm of social interaction, where individuals function as processing units embedded within a complex system (i.e. the ensemble) whose goal entails aesthetic and emotional communication. Each unit possesses the capability to transmit sensory information non-verbally, as well to decode other's movement via the mirror matching system. As information flows along these two channels simultaneously, each unit—and the system as a whole—relies upon predictive models to meet the real-time demands of interpersonal coordination. Thus, the musical ensemble constitutes a dynamical system that possesses important constraints due to the fact that co-performers behave in a complex but formalized (rule-based) manner dictated by musical conventions and often a notated score. We argue that these constraints are beneficial from an experimental perspective. The score played by the musicians, the inherently rewarding experience of achieving interpersonal synchrony, and the intuitive and natural form of social interaction collectively translate into ready-made experimental tasks, high levels of intrinsic motivation, and rich ecological settings.

2.2 The case of Orchestras and Quartets

In a series of studies, sensorimotor communication was measured via the Granger algorithm, applied to movement orchestras and quartets kinematics. The Granger methodology tests whether knowledge of one signal significantly increase prediction accuracy of the future state of another signal ((Granger 1969)). In other terms, the algorithm measures whether one musician's behavior has any influence on the behavior of another musician.

The first study specifically focused on the communication dynamics between orchestra conductors and the musicians (D'Ausilio et al. 2012). It was described that the complex pattern of sensorimotor communication in the orchestra scenario as well as the effect of such interaction on the perceived quality of the musical output. It was shown that the driving-force strengths towards musicians and communication strength among players was modulated across pieces. Interestingly, when the increased influence of the conductor was paralleled by a significant reduction in inter-musician communication, the musical piece was better rated by independent professional musicians. Rather, when the conductor exerted an increased drive, while this was not paralleled by a reduction in inter-musician influences, resulted in worse musical performance. It was concluded that a simple increase in conduction drive might be detrimental to perceived quality if this is not followed by a reduced inter-musician interaction.

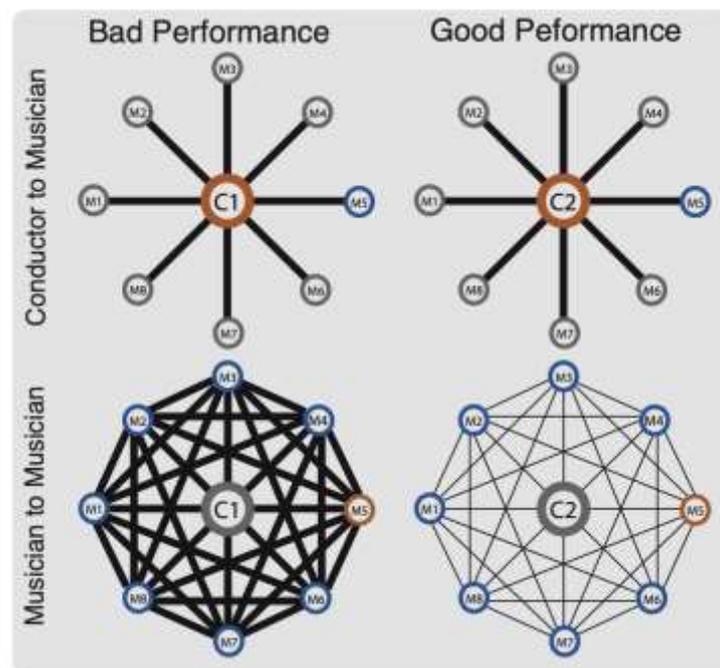


Figure 2. This figure shows a graphical description of the main results obtained by in the 2012 study (D'Ausilio et al. 2012). Here you can see the two indexes derived by the application of the Granger algorithm to movement kinematics – “Conductor to Musicians” and “Musicians to Musicians” – and quality of performance. Quality of performance was provided by expert musician on several dimensions including technical and aesthetic aspects. Results shows that good performance was associated to large “Conductor to Musicians” drive paralleled by a reduction in “Musicians to Musicians” communication.

The second study, on the behavioral coordination in quartets, showed that information flow between all musicians as a group, was larger during segments characterized by technical coordination difficulty. These effects corroborate the hypothesis that musicians coordinate behavior by maximizing their coordinative efforts in specific and critical moments in time (as in the coordination process described by Clayton et al., 2019 and submitted). Furthermore, was shown the emerging of an implicit hierarchy between musicians, with the first violin being most frequently the leader of the group. Then, by applying a perturbation to the communication pattern between the first violin and the rest of the quartet, it was evidenced a specific change in group sensorimotor communication. Inter-musician communication increased whereas the driving force originated from the first violin (the only one knowing when and what alteration to perform) decreased.

The perturbation forces an artificial and temporally confined change in the group dynamics. Such an approach, although it introduces a rather unnatural condition to the normal musical performance is the only reliable method to infer a relation between our Granger-derived indexes and musicians' behavior. Specifically, the increase in inter-musician communication was probably a result of increased uncertainty and thus the need for greater information transfer. On the other hand, behavior of the first violin was particularly interesting since he did reduce his drive towards other musicians. This result can be interpreted in the framework of action anticipation abilities. In fact, quartet musicians can derive coordinative information from the score, as well as from models of others' behavior. Strong musical expertise, as well as continuous rehearsal as an ensemble, certainly helps in increasing reliability and specificity of these models. In this perturbation approach, suddenly the score information became almost useless and other's behavior more unpredictable.

Therefore, musicians might need to shift from a mainly feed-forward control to a more feed-back based strategy. However, reliance on a feed-back strategy do not permit fast and accurate motor control due to inherent temporal delays in sensory signal (D M Wolpert, Ghahramani, and Jordan 1995)(D M Wolpert and Kawato 1998) and multi-agent coordination (Wolpert, Doya, and Kawato 2003). In fact, the advantage of modeling other's behavior is that we can anticipate and sample incoming sensory information less frequently and mainly to confirm our hypotheses on the environment. Therefore, in this experiment musicians probably had to completely shift their normal manner of communication, possibly "reading" others' behavioral cues in real-time while "sending" informative signals more often. Therefore, a possible interpretation is that the reduction in driving force from the first violin signal a reduced efficacy in communication caused by the abrupt departure from a learned manner of communication.

Behavior complexity, together with almost negligible lags in performance (Kokal and Keysers 2010) suggests that purely reactive mechanisms cannot be effective. Others' action anticipation is a necessary prerequisite for successful joint action control (Knoblich and Jordan 2003)(Pecenka and Keller 2011). Generally speaking, expertise has been shown to support other's action classification by modeling future information earlier (Aglioti et al. 2008) and with greater detail (Calvo-Merino et al. 2005). Our musicians were almost at the top level of musical skills, thus suggesting that a strong degree of anticipation of others' action was taking place.

Beyond global descriptions of musician's pattern of relationships, the complexity of these kinds of scenario could also be exploited to distinguish and evaluate the existence of multiple channels of communication as well as their respective role in efficient coordination. In previous studies, one representative kinematic parameter was used to extract global coordination (D'Ausilio et al. 2012)(Badino et al. 2014). However, we know that movements of different body parts may convey substantially different types of information. For instance, bow movements in violinists directly control the sound output (i.e., instrumental gestures), whereas complementary torso oscillations may serve a secondary communicative purpose (ancillary gestures; (D'Ausilio et al. 2015)). More importantly, movements of different body parts may act as different channels of communication, possibly with different roles depending on the specific communication mode. For example, within a quartet, musicians have specific roles while in orchestras, musicians generally play in distinct sections (e.g. sections of violinists). This means that in the orchestra scenario, different modes of communication coexist: a complementary coordination with the conductor and other musicians, in parallel with an imitative coordination with musicians of the same group (playing the same score).

In a recent study (Hilt et al. 2019), it was demonstrated that the pattern of sensorimotor information carried by two selected movements (head and bow) are distinct. Bow kinematics exhibit a robust leader-follower relationship between the conductor and the two violinists' sections. This pattern is robust and is not affected by the experimental manipulation of the sensorimotor information flow

(perturbed condition) except for a decrease in communication between the first section and the conductor. Perturbation consisted in a 180 degrees' rotation of the first line of violinist, such that the conductor has direct visual contact with the second line only. The fact that the perturbation did not dramatically alter the information exchanged via instrumental movements suggests an important role of memory, score reading and residual sensory cues. Ancillary movements, instead, are supposed to convey slower frequency signals possibly related to the expressive component of musical execution, which is more likely to be affected by perturbation of the interaction dynamics. In fact, in head data, the perturbation produced clear alteration of the communication pattern. Communication between the first section and the conductor or the second section was reduced. At the same time, communication between the second section and the conductor increased in both directions. This global increase suggests a greater need for information exchange during the perturbation. Instead, moving to the relationship between first and second line, a complete reversal of their mutual communication was present. Before the perturbation, the first section provided larger causal drive towards the second, while after, the second section lead the first. During the perturbation, the first section no longer had visual contact with the conductor, significantly reducing his role in leading orchestra dynamics. Although violinists of the second section did not actually change their position, they seem to increase their normal communication with conductors, while at the same time they dramatically change the way the communicate with S1 (Figure 3).

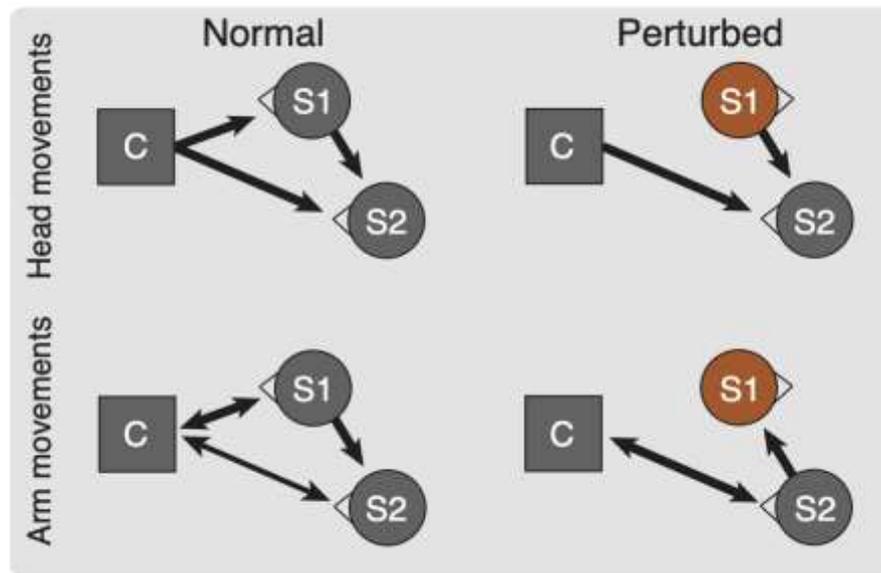


Figure 3. Graphical description of the results obtained in a study investigating hand and arm movements in orchestras.

Here was manipulated the information flow between the conductor, the first and the second line of violinists. The perturbation was a half-turn rotation of the first section, thus facing the second line. Perturbation to this system produces a reconfiguration of the sensorimotor communication network leading to a new centrality of the second line. Interestingly, effects were mostly limited to ancillary movements, which are supposed to convey the expressive component of musical performance.

In conclusions, the present studies show independent confirmations that Granger methods on movement kinematics measure information flow between complex multi-agent interactions. Measuring the temporal deployment of sensori-motor communication between several individuals, in a realistic scenario, opens to a series of important applications. In fact, current communication analyses hardly

offer quantitative results when dealing with complex natural situations. At the same time, this data hint to the fact that efficient sensori-motor communication, may not reside in the unspecific increase in information transfer. One intriguing possibility we suggest is that the fabric of expert interaction is not the ability to communicate per se, but rather the ability to modulate such information transfer when this is more necessary.

2.3 Sensorimotor communication

Concluding this section, successful human-to-human interaction requires important behavioral adaptation, as well as prediction. A large body of literature has focused on cooperation towards shared goals, where humans must combine available sensory information with internal movement production models. In this regard, researchers investigated how dyads achieve interpersonal simple sensorimotor coordination, such as walking side-by-side (van Ulzen et al. 2008) or rocking in rocking-chairs (Richardson et al. 2007). In such contexts, co-actors continuously influence each other and tend to spatially and temporally synchronize their movements. Beside imitation, action complementarity plays a key role in inter-individual coordination with the goal of achieving efficient collaboration (Newman-Norlund et al. 2007). Social interaction indeed goes beyond synchronization with other's actions and relies also on inferring others' motor goals and intentions to generate a context-appropriate action. To achieve fast inter-individual coordination, individuals may build internal predictive models of other's behavior. In function of the context, the most appropriate motor model is compared with the current observed movement, to generate a prediction error (Friston, Mattout, and Kilner 2011) and update own motor planning (Sebanz, Bekkering, and Knoblich 2006).

Due to the technical and analytical complexity in exploring the details of human sensorimotor interaction, only few experiments went further than a dyadic set-up (Dikker et al. 2017)(Alderisio, Bardy, and di Bernardo 2016)(Codrons et al. 2014). However, in daily life, things are usually much more complex. For instance, during a conversation, information is sampled through multiple channels (e.g. vision, audition), sometimes in parallel (e.g. information in the foreground and information from the background) and at different temporo-spatial scales (e.g. slow whole-body movements versus fast lip motions). At the same time, different kinds of information may be conveyed in parallel through different channels. For example, in speech, bodily gestures and spoken words are generally co-expressive. In this context, communication requires flexible means to integrate multimodal data, across multiple timescales and act accordingly. Therefore, proper quantification of (realistic) group coordination is today one of the key missing elements to understand how humans manage to interact with others by efficiently selecting, processing and sending information. In this project, in particular we are interested in content of the information exchanged through movement that conveys emotional qualities.

3. MOVEMENT MARKERS OF EMOTIONAL QUALITIES

Neurologically healthy humans can detect emotions from other people and in the art (films and photographs, theatre, dance or music) through perception and cognition. This capacity is rooted into the detection of signals embedded in static and dynamic patterns of emotional expressions and their comparison to others originating from their previous experience; innate knowledge plays a role as well

(Matsumoto and Willingham 2009). A plethora of research has provided different models of emotions (e.g.,(Ekman and Friesen 2013);(Izard 1971);(Plutchik 1962);(Russell and Dols 1997)) and have evidenced that agents can pick up emotions from facial expressions (e.g., (Darwin 1872),(Ekman, Levenson, and Friesen 1983);(Cordaro et al. 2018)) or from the way people walk (e.g.,(Roether et al. 2009)). In EnTimeMent we embark on the scientific quest to use motion recordings (MoCap recordings) as a primary source of information about one's affective state through Machine Learning Algorithms (MLA) decomposing the signal and clustering it to match capabilities of detection pertinent to the human perceptual system. At the current state-of-the-art, this task is still more challenging for computer science than for human cognition as MLA usually processes data from one source (i.e. visual data for computer vision) and the context information is lacking (information what is happening in the environment derived from other signals that could inform about emotional state of the agent). Our approach to this complex undertaking is built on the several assumptions, derived from the existing body of research suggesting that i.) movement qualities (i.e. speed, amplitude) are directly steered by emotional arousal; ii.) movement is a potent messenger of one's emotional state for the environment ; iii.) emotional expression depends on the context (i.e. could be diminished is in a non-familiar setting); iv.) different timescales might be employed for different emotional qualities.

Emotional expression enriches communication with others and enables to react in a timely manner to the changes in the environment. All animals are programmed by nature to move to obtain food (provide energy), avoid danger (survive) and reproduce (pass on genetic information). From a biological perspective primary function of emotion is to make the interactions with others more efficient (for example to co-operate), avoid threats and thrive in the surroundings ((Barrett et al. 2019)). In the same vein, emotions can be understood as modes of action readiness relevant to the context a person is in (Jackendoff 2009). The outcome emotion is derivative to the valence of the interaction (positive, negative) and level of arousal (low or high) (Kuppens et al. 2013) modulating the agent's drive to approach or withdraw from source of emotion (object or person). Neuroimaging studies show that sensory stimuli (acoustic), which is emotionally loaded, modulates the excitability of the corticospinal motor tract (Komeilipoor et al. 2013). For example, in simple terms, fear promotes readiness for fight or flight reaction via high arousal (readiness to spend high energy) with negative valence (avoidance). At this point it is essential to note that the emotional state is personal, highly unpredictable and differs from one individual to another (i.e only 9% of emotional expression consistency was noted for facial expression of fear, (Barrett et al. 2019)). Nevertheless, at least some features of emotional expression seem to have shared properties for humans ((Ekman and Keltner 1997); (Frijda and Tcherkassof 1997)), and could be universally recognized by other people without prior personal knowledge. For example, previous research has found that for example anger and happiness movements are faster and jerkier than movements associated with sadness, which are slower and smoother (Pollick et al. 2001).

In EnTimeMent we aim to determine whether it is possible to detect emotional states from the motion derived data with and without any known priors, which could help us to identify how we are going to extract that information in the signal. Since the pioneering work of Gunnar Johansson ((Johansson 1973)), it is known that motion signals are potentate carriers of information about individual characteristics ((Loula et al. 2005); (Cutting and Kozlowski 1977); (Troje and Westhoff 2006)), but little is known about quantifying emotional qualities from movement data. Importantly, some emotions – such as fear and disgust – were found to be harder to identify than others ((Dahl and Friberg 2007); (Atkinson, Tunstall, and Dittrich 2007)). From the other side, more recently it was pointed out that the neutral state, while frequently ignored within the domain of emotion studies, is as important as any other emotional state and should be incorporated to any emotion model (Sacharin, Schlegel, and Scherer 2012).

Music performance can again offer a useful testbed, since in some genres at least performers' ancillary movements are linked to emotional expression. For instance, in North Indian raga performance different modes (ragas) are linked to particular moods, images, and/or emotional expressions; while performing them, singers appear to instinctively use distinctive movement patterns in their manual gesture. One of the areas to be explored by EnTimeMent is whether elements of performer movement linked to particular emotional expressions can be reliably identified.

3.1 Introduction to IMS and IEMS

As a first step to set the background for the Machine Learning processing of the data, we need to define and detect neutral state which corresponds from the informational point of view to a maximum entropy state. This means that when a person is in its neutral state, it is impossible to detect any emotional quality from the motion data because he or she does not experience any emotion providing a fruitful source of information serving as a baseline comparison, thus, an increasing number of scholars incorporate the neutral state to their experimental paradigm (e.g. (Kleinsmith and Bianchi-Berthouze 2013)). This neutral state will be collected for each individual as a baseline reference prior to the inducing other emotional states through manipulation in the laboratory. We will further transform this data to Individual Movement Signature (IMS) building up on the work developed in EuroMov during ALTEREGO project as described by Słowiński et al. (Słowiński et al. 2016) and Coste et al. (Coste, Bardy, and Marin 2019).

To be able to extract emotional states of the individual data we will use his/her neutral state as a baseline for drawing further comparisons (Vergotte et al. 2018). All basic emotional states (e.g., happiness, sadness, anger, fear – insecurity, or disgust) will be evoked via enactment, deception or emotional induction with sensory stimuli during a task involving improvised movement (a possible scenario). A plethora of research suggests that emotional content is best conveyed by the distal parts of the body (with relatively low energy expense cost for moving) during gross motor acts with high number of degrees of freedom (Roether et al. 2009). Based on those assumptions, we will search for the invariant movement features that we refer to as Individual Emotional Movement Signatures (IEMS) derived from the observation of change of IMS during emotional stimulation experimental conditions.

Upon completion of this step our scientific ambition is to answer the question whether any components of IEMS are shared amongst individuals. In other words, we will search for interindividual similar patterns in the dimensions of space and time. Indeed, the similarity measures are the cornerstone of most data science methods with plethora of options available ((Harispe et al. 2017); (Tami et al. 2018)). We will search for the best fitting similarity measure to compare IEMS across different individuals. Another approach is to search for the least quantity of information that describes the dissimilarities. This information is encapsulated in the concept of signature. Searching for dissimilarities is challenging as this notion does not share the same mathematical properties as similarities do. Nevertheless, recent MLA techniques such as sparse learning (Jiu et al. 2019) help to locate a minimum set of features that describe a given phenomenon.

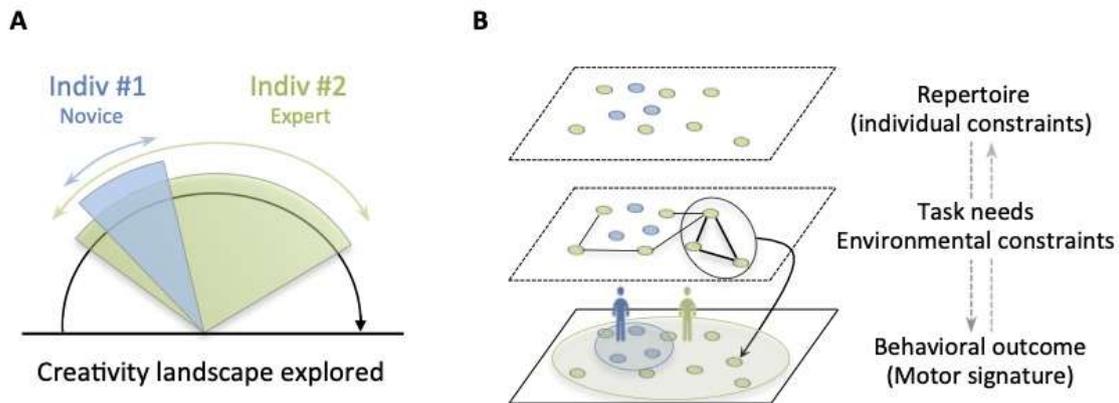


Figure 4. (A) Schematic depiction of the motor creativity of two individuals with different levels of expertise in improvisation. The landscape of creativity explored by the expert (indiv#2) is supposed to be broader than that of the novice (indiv#1). (B) Illustration of the three layers leading to the emergence of improvised behaviors. The first layer corresponds to the individual repertoire formed largely by past body activity experiences. The second layer shows how behavioral solutions emerge through the exploration of the individual repertoire under external constraints (task and environment). The third layer corresponds to behavioral outcome - the observable part of the improvisation process at the behavioral level - which can be easily captured by means of movement analysis and made readable using dimension reduction techniques (e.g., multidimensional scaling to display the individual motor signatures). In this way, each individual's trial was plotted as a point on a map, so that similar trials are placed near each other and dissimilar trials are placed far from each other. The area of ellipses that encompasses all dots (experimental trials) of each individual provides a measure of within-person motor (intra-individual) variability (i.e., creativity landscape explored) (Coste, Bardy, and Marin 2019)

3.2 Multi-Time mode Concept

Finding common IEMS patterns among individuals would support hypothesis that emotional expression during movement is governed by universal, neurological principles for humans. Evidence provided by extended EEG imagining of the brain (LORETA method) with passive affective stimuli suggests that all emotional stimuli is processed by similar neural network (Costa et al. 2014). However, it is the temporal aspects of the processing that were found to change depending on emotional quality, with the hypothesis being that each emotion has its own neural signature in the brain. The differences were noted in rate of the spike of the neural activity to the gradual rise and decrease for emotions having more profound and personal meaning (happiness or sadness, leading to long-term change in behavior; (Frijda and Tcherkassof 1997)) and faster and shorter for reactive (emotions needing immediate interaction with the environment) with the processing time for the affective stimuli being below 500ms in all cases. Based on the current body of literature encompassing both neuroimaging and cognitive sciences it is impossible to define how long the emotional expression lasts on average in humans. Important to note here is that emotions serve interaction with environment therefore their lifetime also depends on the circumstances of situation and changing context (i.e. whether threat was withdrawn from the surroundings of individual or not), which can cause high inter- and intra- individual variability. In a self-report study on the group of 230 Dutch students some emotions were reported to last from minutes (anger, shame), hours (joy) to days (sadness)(Verduyn et al. 2015). Such multi-timescale hierarchical organisation is a well-known property of all animal world

with fast dynamics being characteristic for lower hierarchy movements and slower ones for upper hierarchy ones, which allows flexibility of behavior (Kaplan et al. 2019), in humans referred to as “motor grammar” (Bernstein 1967). In this view, the motor system can recombine or substitute motor elements to cope with changes in the context giving rise to different qualities of movement. Therefore, humans behaviour naturally occurs in parallel at different timescales, also making the experience of time more subjective in human perception. For example; humans naturally produce rhythmical movements and perceive periodic signals in range between 100ms and 1000ms, with longer timescales stretching from 1 second to 5 seconds characterising action sequences or other processes such as heartbeat or winking. Perception of movements over course of minutes allows agents to interpret meaning and relationship between different action steps. It takes hours to learn a simple movement and days to learn a complex skill. Over years we age, change and adopt more functional action-perception patterns and here the range of seconds (duration and temporal resolution wise) is lacking saliency in meaning.

This gives rationale to hypothesize that IEMS will be identified across timescales of different duration and resolution depending on the emotional quality stamp sought in the MoCap data. Therefore, in each experimental design we will focus on capturing at least two different groups of emotional qualities in comparison to the neutral state. The neutral state, baseline IEMS, will be applied to differentiate whether the same timescales are relevant or need to be adapted to each emotional quality. For example, it is almost intuitive to assume that sadness related qualities will emerge with lower frequency of the movement. This would imply stretching the MoCap period to longer durations, based on the evidence from other research showing that velocity is one of the key differential variables for recognizing the emotional qualities in the movement ((Pollick et al. 2001); (Sawada, Suda, and Ishii 2003)). In order to use supervised MLA all observations will be self-annotated or annotated by external observer(s) at each stage of the development.

4. MACHINE LEARNING FOR MOVEMENT ANALYSIS

In EnTimeMent the scope is to propose a radical change of paradigm and technology in human movement analysis, where the time frame for analysis is grounded on novel neuroscientific, biomechanical, psychological, and computational evidence, and dynamically adapted to the human time governing the phenomena under investigation. In order to reach this goal it is necessary to propose and develop innovative scientifically-grounded and time-adaptive data analytics and machine learning technologies that can operate at multiple time scales in a possibly multi-layered structure. The scope is to exploit the information captured by the current generation of motion capture and movement analysis systems and empower them, with these needed computational models, to achieve a novel generation of time-aware multisensory motion perception and prediction systems. In order to reach this goal, different data analytics and machine learning technologies can be employed. In particular, EnTimeMent will investigate two main families of methods (see Figure 5).

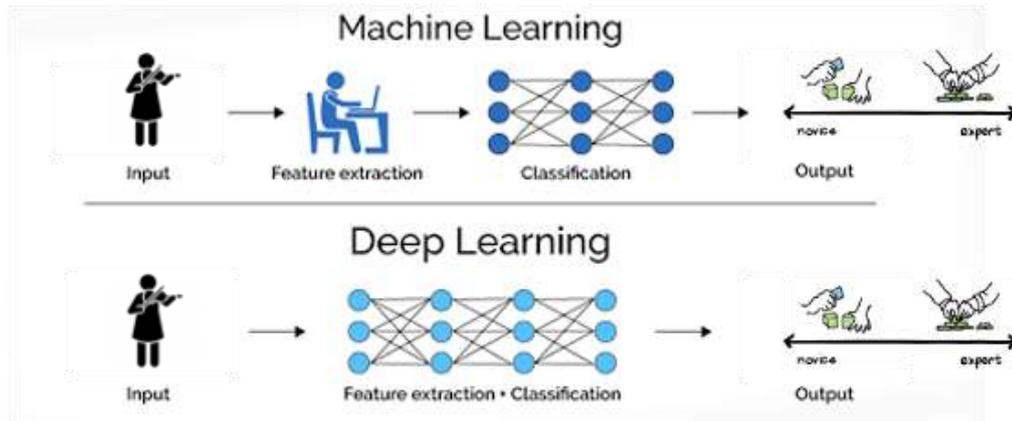


Figure 5. Schematic depiction of the different families of computational methods employed in EnTimeMent.

The first family considered are the classical Machine Learning models (Ensemble Methods, Kernel Based Methods, Shallow Neural Networks, Bayesian Methods, etc.) (Shalev-Shwartz, Ben-David, 2014) where features are extracted from data based on the knowledge of the problem, during the feature extraction phase, in order to create a rich and expressive description keeping in mind that the scope of EnTimeMent is not just to create models able to detect and characterize the body movements and but also to show that this phenomena has multiple time scales that needs to be detected in order to well characterize them. In order to reach this goal features must be engineered so to capture information at multiple resolutions of time and space.

The second family of methods, called deep learning models (Goodfellow, Bengio, Courville, 2016), removes the features engineering phase and replaces it with another Machine Learning model, usually a (deep)-neural network, able to automatically learn the rich and expressive representation of the data automatically from the data itself. In order to reach this goal a very large variety of methods and architecture exists but, for the purpose of EnTimeMent, Recurrent Neural Networks are surely the most suited architecture able to automatically learn a rich and expressive representation able to express and describe the movement at multiple time scales. In fact, RNN can, with different techniques (Long Short Term Methods, Clockwork Networks, Attention Mechanism, Convolution in Time, etc) are able, in principle, to capture the multiple time scales of a phenomena.

With respect to the first family the deep learning models, in general, require much more data to be trained and are more prone to overfitting. For this reason it is always necessary to exploit both families of methods to ensure the good quality of the data, the meaningfulness of the extracted information, avoiding to capture spurious correlations, the statistical robustness and consistency of the methodology. In this sense, rigorous statistical procedures for model selection and error estimation (Oneto, 2020) together with a posteriori analysis of the learned models with feature ranking/selection, attention maps or, more generally, some techniques to make model explainable (Molnar, 2019), should be employed.

4.1 Discovering relevant movement patterns

Traditional methods for human movement analysis have relied heavily on feature engineering, using predefined features that have shown to be indicative of the quality of movement. However, with

engineered features, information deemed to be irrelevant by the designer is discarded from further consideration, leading to a potential bias in the analysis. The analysis cannot say more than what is contained in the features, which are in turn based on what the designer believes to be relevant. It is for example increasingly clearer that psychology frameworks (e.g. FACs) used to inform the design of features for automatic facial expressions recognition are limited in capturing everyday facial expressions (Barret et al., 2017). More relevant to EnTimeMent is the work by Avizier et al. (2012) showing that people are convinced to make most of their evaluations of other people's emotions on the basis of their facial expressions while often they are using features of the body movement to reach their conclusions. With the introduction of modern representation learning methods, typically based on deep neural networks, such bias can be avoided (Bengio et al., 2013). Instead of engineering features, features are learned from the raw sensory data with the influence of the designer kept to a minimum.

Features that emerge during the learning process may or may not resemble those of earlier designed features. Regardless of which, the learned features may be of interest for further investigation. Some features may be used to confirm the relevance of patterns that are believed to be indicative of particular movement qualities. It is also possible for new relevant patterns to be discovered, in particular patterns that are hard for a designer to conceptualize and define as a feature. If the learning framework allows for signals to be represented over multiple scales, the temporal scale of the pattern can potentially also be discovered.

Unfortunately, most learning frameworks do not explicitly indicate the movement patterns that are most indicative of a particular prediction. Notable exceptions are attention-based methods that apply weights to place more or less emphasis on the data given how important they are assumed to be for the prediction (Liu et al, 2017). For feed-forward neural networks, a common approach to understand the inner workings of the predictor is to propagate back gradients from a predicted concept to the most salient part of the input data, resulting in a heatmap that can easily be visualized ((Selvaraju et al. 2017), (Stergiou et al., 2019)). It should be noted though that much is still unknown what deep networks actually learn and what leads to a particular prediction. This can be illustrated by their lack of resistance towards so-called adversarial attacks (Akhtar and Mian, 2018), where small unnoticeable changes to the input can significantly alter the prediction.

4.2 Predicting multiple futures

Another application of modern machine learning methods is to predict future movements. Given the recent history of movements and qualities derived from it, it can be predicted how an agent will move in the near future. If movements are modeled in a probabilistic fashion, future predictions can be described as distributions that capture multiple possible futures. Even in cases where only the most probable future will eventually be used, the variance of the distribution can be viewed as a measure of the uncertainty of the prediction.

There could be many reasons why a system would benefit from future predictions. You might have multiple interacting agents, where predictions are used for planning for an agent to move in accordance with the movements of others. A prediction can also be used as a means to verify whether the system's understanding of the movement is likely to be correct by comparing it to the true future movement. Another possible reason is to detect events that suggest a change in state when the agent changes from one type of movement to another.

Most research of movement prediction has so far been done on relatively short time scales, typically not more than a few seconds. This is true at least if predictions are given in the space of the original input data ((Fragkiadaki et al., 2015), (Bütepage et al., 2017)). Most such frameworks rely on some kind of encoder-decoder structure, where the encoder compresses recent movement into a low-

dimensional representation, which is then used by the decoder to predict multiple futures. For human-robot interaction, the decoder can often be used as a generator to generate appropriate movements in response to the movement of a partner. In the context of physical rehabilitation (one of the cases of EnTimeMent), such decoders could be used to drive real-time generation of movement sonification to enhance movement awareness or even help correct movement trajectory before they take place by using embodied mechanisms of sound (Newbold et al., 2016).

For longer time scales, predictions are typically given in terms of some higher-level movement qualities, such as the activity the agent is currently engaged in. To reduce the complexity of the learning framework that would otherwise grow, as longer time scales are used and more data is processed for prediction, learning is often done in stages. Data are first represented as streams of learned spatio-temporal movement features that are later aggregated for prediction ((Ryoo, 2011), (Li and Fu, 2014)). Since this division into two stages might lead to bias by design, there is a trend towards end-to-end training, where features and predictors are learned at the same time (Kong et al., 2020).

4.3 Abstracting representations of movements

Regardless of which machine learning method you use, over-training easily becomes a critical concern as the temporal scales from which predictions are made increase. Learning frameworks increase in complexity and with more unknown free parameters to train, the need for more training data increases far beyond what is feasible to capture within a reasonable time. Without a sufficient amount of training data, the excess learning capacity of frameworks will instead be used to learn particular patterns in the training data that are of no real significance, making the frameworks unable to generalize beyond the training data. A common solution in computer vision and machine learning is to pre-train networks for an auxiliary task for which large amounts of data are already available. This is not always the case though. Annotated training data with human participants is always expensive to acquire in particular if it involves partners that are interacting or data from clinical populations.

Fortunately, the repetitive and cycling patterns of movement data tend to make the data highly redundant. If the training data include annotations of movement qualities to predict for a particular task, information redundant for this task can be eliminated by training the framework in a supervised setting, which in turn reduces the complexity of the predictor and the need for large amounts of data. This does, however, prevent representations from being used for other tasks, at least as long as the two tasks are different enough with respect to the nature of the information required by their respective predictors. An alternative is to use unsupervised training and reduce the dimensionality of data while representing it in a form in which more information is preserved.

Given the nature of movement data, it comes as no surprise that traditional methods for dimensionality reduction have been applied to reduce its size, e.g. using wavelet transforms (Beaudoin et al., 2007). The size is reduced by discarding dimensions that have the least impact on a reconstruction from further analysis. However, whether data is redundant or not depends on the application. Dimensions might be removed despite being essential for the prediction of particular movement qualities if these dimensions have little effect on the overall movement. It thus becomes important to reduce the complexity of data, while at the same time ensuring that information necessary for prediction remains. Another way to compress the data is to learn some kind of sparse representation of the data using primitives known to be relevant for the target application. These primitive can be localized in time and space modeling individual joints or markers ((Schaal, 2006), (Li et al., 2010)) or could be a more global temporal segmentation of the movement (Lu and Ferrier, 2004). However, there is a risk for

introducing a bias by design given by the limitations of the approach chosen, in particular for the temporal scale during which significant events are expected to occur.

More modern methods in machine learning try to limit the influence the human designer has on what is eventually learned. Many such methods are based on variational autoencoders, feed-forward neural networks that are trained to predict what it has on its input but does so through a narrow waist represented by a layer consisting of a small set of neurons. Once the network has been trained, the activations of these neurons can be regarded as a low-dimensional latent representation of the original data (Kingma and Welling, 2014). These autoencoders are flexible in the sense that they can be easily extended with additional objectives related to the particular task, allowing e.g. ground truth data of known movement qualities to affect the way the latent representation is structured through semi-supervised learning (Bütepage et al., 2018). By doing so representation can remain compact, while still keeping dimensions of the data known to be important for prediction.

4.4 Coupling and transfer of information

Another benefit of probabilistic modeling, possibly using variational autoencoders or similar networks, is to model the coupling between different units. These units could e.g. be different moving agents, different tasks or sensor modalities. For example, in (Bütepage et al., 2019) it was tested whether observations of human partners engaged in interaction, different types of hand-shakes in this case, could help a robot to learn similar movement patterns. Learned probabilistic models were used to describe the movement of each individual participant, but these models were conditioned on a latent space representation of the joint task that was also learned. In experiments, it was shown that the robot benefits from such a coupling between models when trying to learn a new behavior, which is true even if the coupling was learned from observations only. This suggests that observations can help an agent learn to engage in an activity, even with the observations do not include the agent itself, but two other interacting agents.

In recent years, transfer learning has become an important tool to allow transfer knowledge from one domain to another (Cook et al., 2013), typically from a domain for which you have a sufficient amount of data to a domain in which it is easier to draw conclusions on relevant movement qualities. One such possibility is to transfer knowledge from video data to motion capture data ((Mehta et al., 2017), (Zhou et al., 2017)) which may have direct practical application in e.g. rehabilitation, which is a target area of EnTimeMent. In addition, as rehabilitation moves from exercise-based sessions into functional activities, there is also the need to transfer to a personalized version of wearable (possibly reduced) mocap technology. Other applications for which transfer learning can be beneficial is to transfer observed qualities of movement into a form that is more easily interpreted by human users, using e.g. sonification.

4.5 Providing insight into the data available and the need for more data

The sections above have discussed some of the possibilities the new machine learning techniques offer to the study and modeling of body movement and body movement qualities at different time scales. The application of such machine learning techniques to available human activity and affective body expressions datasets can already help shed more understanding on important movement patterns, beyond the one available from the psychology literature (e.g., for a survey see Kleinsmith et al., 2013; Karg et al., 2013) to better understand how emotion is expressed through the body by comparing and

modeling datasets from multiple contexts rather than independent single individual datasets and what movement parts are cues to understand people interaction dynamics.

Unfortunately, motion-capture-based body movement datasets are still very sparse and quite limited to very simple activities. They often include a very small number of sensors and they are generally unimodal (see Deliverable 1.1 for a survey). This is even more the case when we consider affective qualities of movement rather than activities. Most of these datasets consist of acted rather than naturalistic data. This sparsity of data limits the possibility to apply more advanced machine learning techniques and even to develop new algorithms that address the specific characteristics of such data and of the related modeling applications.

A specific need emerges from the fact that affective body expressions do not happen outside a context and such context contributes to shape them (Barret, 2017). A particular context is the activity the person is doing. To fully understand the affective quality of body expressions, there is a need to gather those within the context of a large variety of activities rather than isolation. The trigger of an emotional response and the environmental and cultural context (what is socially acceptable) the person is in are also factors that contribute to modify the expressions. This variability is critical as affect recognition and activity recognition systems are starting to be deployed in specific real-life applications beyond entertainment and research. While voice and face are the main modality industry is currently considering, we can already see how assumptions are made by industry on the generalization capabilities of this technology. However, facial expression recognition can rely on advances made in computer vision and on the large datasets that are now being collected using convenient front-of-camera settings (e.g., laptop). The work on affective body expressions suffers from the fact that such settings do not easily fit full-body activities (general just face and shoulder) and as said above HAR is also under-developed as often these datasets are privy of emotional responses as data are conducted in quite repetitive sessions.

A number of full-body datasets have been created in the contexts of dance, music playing, and computer games. Again these are mainly sparse activities and there has been very little attempt to understand how these resources can be brought together and exploited. This is an activity that EnTimeMent may aim to address given that many of the available datasets have been developed by the consortium partners. In doing so, EnTimeMent may attempt to set guidelines for such recording, sharing and building a benchmark dataset platform.

A domain that is even more critical to cover is clinical datasets ((Lucy et al., 2013), (Riva et al., 2018), (Aung et al., 2016)). These are critical for building effective applications, however, they are very much time consuming and suffer from the fact that, unless built in the hospital (e.g., Joshi et al., 2013), they may be biased toward a clinical population that is able to cope with the physical and psychological demand of such dataset. For example, outside of the clinic data collection, may contain less people that suffer from depression as such a population may be more reticent to engage in a psychologically demanding social activity of a typical data collection session. To address such limitations (among giving also the possibility to gather more real-life data and useful data), in EnTimeMent, we are aiming to collect data of people with clinical conditions (chronic pain) in their own homes. This raises technical challenges to data collection as complete and more reliable sensors architecture cannot be used. In such settings, for example wearable with a limited number of sensors (to address acceptability, comfort and reduce sensor interference) can only be used (Olugbade et al., 2018).

Another important point to consider is the need to go beyond simple movement sensors to capture the skeleton of the body. EnTimeMent sees movement as a more complex process. Muscle activity, respiration or even neural activation patterns provide different information about movement and about its affective drives. Such rich platforms create interesting new questions for machine learning (their different role, off-set, temporal scale, etc.).

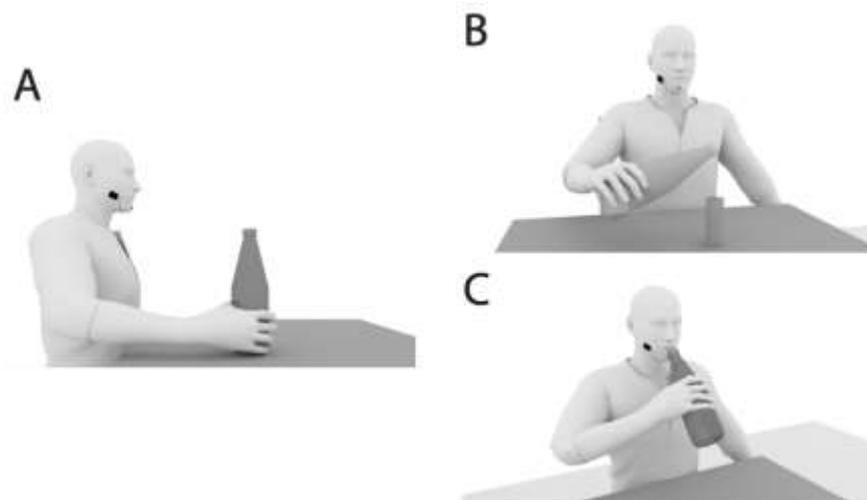
Whilst all movement datasets can be said to be built mainly for very short temporal scales in mind, wearable devices could offer new opportunities in terms of long term temporal scales. Such data would allow the study of movement perception and modeling at a very different level (e.g., habits, perception of capabilities, long term prediction). A proxy to such data collection can be seen in a multi-session data collection for the same person. Such datasets are still very limited and mainly considered only in stationary situations or for a very limited number of sensors (e.g., accelerometer in a smartphone).

5. INTRODUCTION TO EXPERIMENTAL WORK

5.1 Individual Level

We propose to develop this approach via set of naturalistic and enacted experiments encompassing both upper limb movements (i.e. reaching for objects of different emotional valence, ball dribbling, pendulum manipulation, musical instrument playing and orchestra leading) and gross body movements (walking, dancing, bouncing, walking up the stairs) to explore unconstrained and improvised movement scenarios.

There are three primary sources for extracting the information from body movement, which are: kinematics, such as velocity and acceleration, for example in respect to the gender, Mather and Murdoch ((Mather and Murdoch 1994)) showed that the ratio hips – shoulders width is not as predictive as the body sway; static form and motion information, such as a body parts configuration where the form changes decrease the accuracy in a greater extent than the kinematic features ((Atkinson, Tunstall, and Dittrich 2007)); dynamics, such as mass and force, as for instance in the ground reaction forces feature extractions of emotions from gait (e.g.,(Janssen et al. 2008)). In the light of our research interest we will investigate the kinematic source. Within the kinematic research, it was demonstrated that the marker placed on the trunk has the best accuracy in discriminating emotions ((de Meijer 1989)). In a more recent evidence, the full the hand movement were discriminative of high activation emotions ((Gross, Crane, and Fredrickson 2010)). Moreover, some scholars have even restricted their research focus to expressive emotional manifestations of arm movements only (e.g.,(Pollick et al. 2001);(Sawada, Suda, and Ishii 2003)). Therefore, inspired from literature our main focus will be on the kinematic parameters of markers of trunk and arms, having in mind the gender differences, but keeping track of the other articulations to obtain a holistic picture. That means the movements recorded should be unconstrained, improvised and involve gross body actions. In addition, a sound signal from the movement produced (i.e. recording of steps on the floor) which can enhance information about movement dynamics ((Young, Rodger, and Craig 2013)) and intentions ((Camponogara et al. 2017)). In the case of music performance research described above, the combination of movement information extracted from the unconstrained movements of performers with a melodic sound trace comprises a rich data set. These signals could be further enriched with breathing recordings to provide more sources of information for the machine learning algorithms to identify IEMS across different conditions. At EuroMov we conducted substantial research during BEAT HEALTH project on the ratio between dynamics of breathing and locomotion systems during various movements ((Bardy et al. 2015)) and shared breathing tempo in group synchronization ((Codrons et al. 2014)). In addition, literature points towards strong links between respiration and emotional arousal in humans and animal world ((Homma and Masaoka 2008)). Taken together it suggests that breathing could be a valuable source of information ((Lussu et al. 2019)) complimentary to the MoCap and sound recordings of movement.



Example of an individual manipulative sequential action. The action involves a series of distinct motor acts (e.g., reaching toward, grasping, lifting, transporting the bottle; A) in order to accomplish some overarching goal (e.g., drinking or drinking; B or C). Chaining of motor acts predicts that electromyographic (EMG) components that are required for the final act (e.g., activating the mouth-opening mylohyoid muscle, MH, for drinking) show anticipatory activation. This prediction has been confirmed in a study measuring MH activity during the execution of grasp-to-drink and grasp-to-pour actions (Soriano et al. 2018). MH activity selectively anticipated the execution of drinking, being significantly higher for grasp-to-drink actions compared to grasp-to-pour action already at the time of first object contact. Figure 5.

5.2 Dyadic level

Do the emotions diffuse between individuals? If so, how does it happen and can we see it in the way people adapt their movements? The simplest topology we aim to explore is the dyadic one. Our experimental approach will involve scenarios, where one individual (the guide) plays the movement of an emotion and see how, and how long it takes for the second individual (the other experimental participant) to adopt the movement and whether the slave will show the traces of emotion of the guide in their own IMS altering it from neutral state to EIMS. Can a dyad create an Emotional Dyadic Motor Signature – EDMS? How does the emotion expressed by the guide impact the unfolding of the interaction (the game), or does some other movement / emotion emerges as the game goes on? All the information gathered from the individual level will be useful here to finetune the experimental design and shed a light on data science approaches for EIMS and EDMS. The dyadic level will be the basic element for more elaborated group topologies, e.g. a star graph with one guide and many slaves, or a complete graph with all agents being emotionally and perceptually coupled to all other agents.

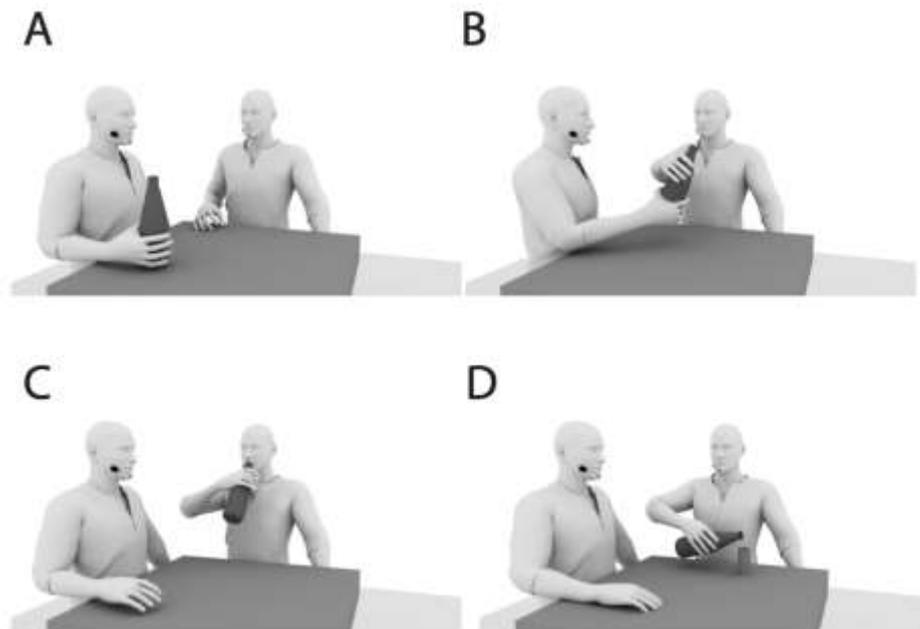


Figure 6. Example of a dyadic manipulative sequential action. The action involves a series of motor acts distributed across two agents. In the schematic, one agent (passer) reaches towards, grasps, and lifts a bottle (A), while the second agent (receiver) reaches to grasp the bottle (B) with the goal of either drinking (C) or pouring (D). Chaining of motor acts leads to two distinctive predictions. First, if chaining extends to the entire dyadic sequence, then EMG components that are required for the final act performed by the receiver should show anticipatory activation in the passer. The distinctive finding would be that already during object grasp (by the passer), MH activity in the passer is greater in anticipation of a grasp-to-drink compared to grasp-to-pour actions. Second, during the same action phase, activity in MH in the receiver should be greater in anticipation of a grasp-to-drink compared to grasp-to-pour actions. This second prediction derives from the finding that the same motor chain organization supporting the execution of sequential actions also support action prediction during the observation of sequential actions ((Soriano et al. 2018)).

5.3 Group level

How does a group emotion reveal into the movement of the individuals of this group? When a group of dancers plays an emotion, their movements are not identical. Are there invariants in these movements that remain the same for all the members of the group? Could one achieve some kind of Emotional Group Motor Signature (EGMS)? We will embark on investigating impact of emotional qualities on group synchronization properties (coupling strength, leadership, topology) with professional dancers enacting emotional qualities in their movement and other scenarios involving group motor activities (i.e. pendulum manipulation). What can one say for the movement of people (in natural conditions) that move spontaneously under the influence of an external event (e.g. the favorite team scores and the supporters jump in the same time). In this scenario, and in addition to controlled laboratory-based experiments manipulating topologies and individual characteristics (e.g., (Alderisio et al. 2017); (Bardy, B. G., Calabrese, C., Bourgeaud, S., Colomer, C., Pla, S., de Lellis, P., & di Bernardo, n.d.)), we are interested to explore recording of natural group events (public transport disruption, audience of football game) with the use of video recordings and deep neural algorithms enabling to track body parts in multiple agents (i.e. DeeperCut or DeepLabCut, (Insafutdinov et al. 2016); (Nath et al. 2018) respectively).

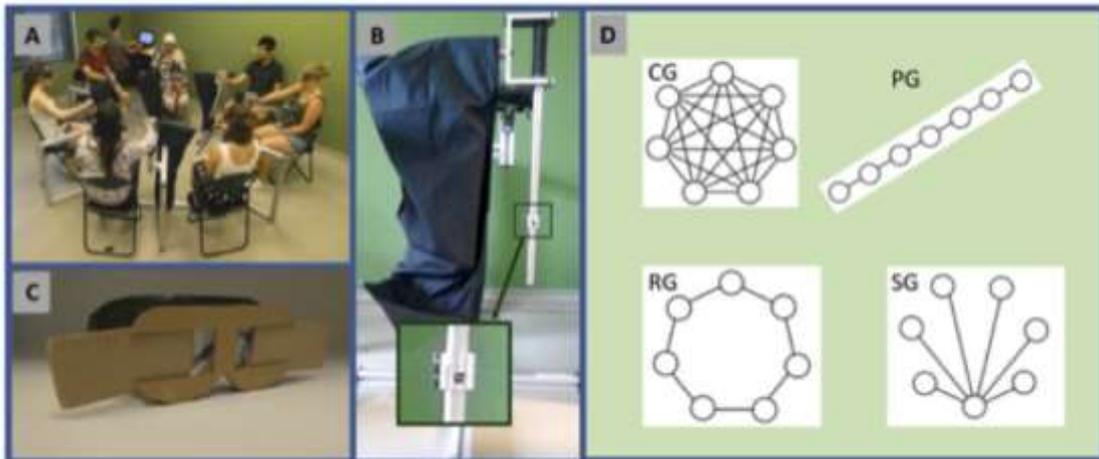


Figure 7. Example of experimental setup exploring the emotional contagion in the group synchronization task. A) participants moving in synchrony, B) pendulum apparatus used for production of arm sway for each individual in the group, C) in-house goggles controlling the field of view; D) complete, ring, path, and star graphs topologies showing common spatial alignment of people performing a motor task with a common goal.

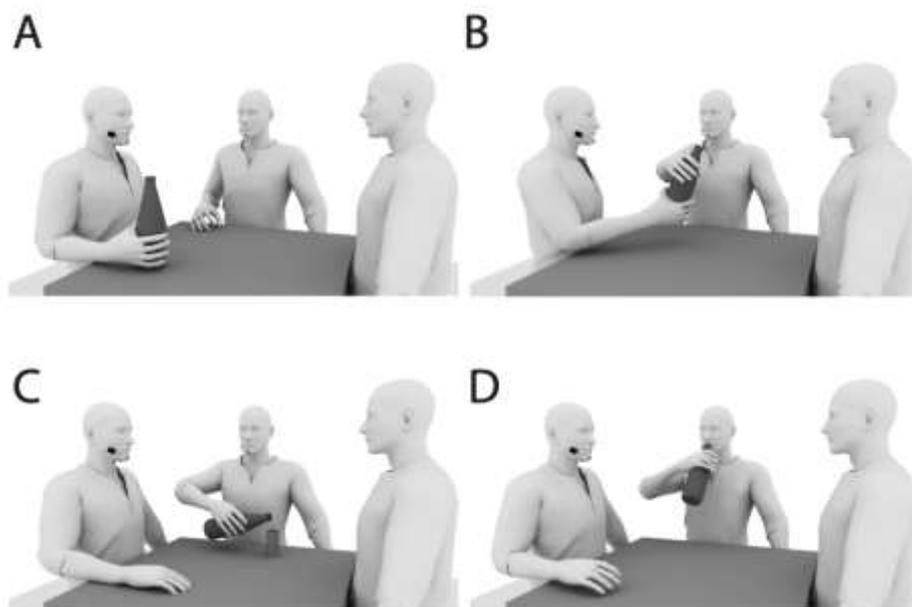


Figure 8. Example of a triadic social interaction. The scenario is similar to the dyadic sequential action scenario (see above), except that, it involves a second receiver (A). The passer passes the bottle to either one receiver or the other (B), who either drinks (D) or pours from the bottle (C). This scenario provides the opportunity to test how participation ((Schilbach et al. 2013)) influences action chaining during sequential manipulation. Based on (Kourtis, Sebanz, and Knoblich 2010) we would expect greater activity in MH in the receiver in anticipation of self-directed compared to other-directed grasp-to-drink actions.

6. REFERENCES

- Aglioti, Salvatore M, Paola Cesari, Michela Romani, and Cosimo Urgesi. 2008. "Action Anticipation and Motor Resonance in Elite Basketball Players." *Nature Neuroscience* 11 (9): 1109–16. <https://doi.org/10.1038/nn.2182>.
- Alaerts, Kaat, Patrice Senot, Stephan P Swinnen, Laila Craighero, Nicole Wenderoth, and Luciano Fadiga. 2010. "Force Requirements of Observed Object Lifting Are Encoded by the Observer's Motor System: A TMS Study." *The European Journal of Neuroscience* 31 (6): 1144–53. <https://doi.org/10.1111/j.1460-9568.2010.07124.x>.
- Alderisio, Francesco, Benoit G Bardy, and Mario di Bernardo. 2016. "Entrainment and Synchronization in Networks of Rayleigh-van Der Pol Oscillators with Diffusive and Haken-Kelso-Bunz Couplings." *Biological Cybernetics* 110 (2–3): 151–69. <https://doi.org/10.1007/s00422-016-0685-7>.
- Alderisio, Francesco, Gianfranco Fiore, Robin N Salesse, Benoît G Bardy, and Mario Di Bernardo. 2017. "Interaction Patterns and Individual Dynamics Shape the Way We Move in Synchrony." *Scientific Reports* 7 (1). <https://doi.org/10.1038/s41598-017-06559-4>.
- Akhtar, N., & Mian, A. (2018). Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access*, 6, 14410-14430.
- Amoruso, Lucia, and Alessandra Finisguerra. 2019. "Low or High-Level Motor Coding? The Role of Stimulus Complexity." *Frontiers in Human Neuroscience* 13: 332. <https://doi.org/10.3389/fnhum.2019.00332>.
- Atkinson, Anthony P, Mary L Tunstall, and Winand H Dittrich. 2007. "Evidence for Distinct Contributions of Form and Motion Information to the Recognition of Emotions from Body Gestures." *Cognition* 104 (1): 59–72. <https://doi.org/10.1016/J.COGNITION.2006.05.005>.
- Avenanti, Alessio, Matteo Candidi, and Cosimo Urgesi. 2013. "Vicarious Motor Activation during Action Perception: Beyond Correlational Evidence." *Frontiers in Human Neuroscience* 7: 185. <https://doi.org/10.3389/fnhum.2013.00185>.
- Aviezer, H., Trope, Y. & Todorov, A. (2012). Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. *Science*, 30, 1225-1229
- Aung, MSH, Kaltwang, S., Romera-Paredes, B. , et al.. (2016). The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal EmoPain dataset. *IEEE Transactions on Affective Computing*, 7, 4, 435-451.
- Babiloni, Claudio, Fabrizio Vecchio, Francesco Infarinato, Paola Buffo, Nicola Marzano, Danilo Spada, Simone Rossi, Ivo Bruni, Paolo M Rossini, and Daniela Perani. 2011. "Simultaneous Recording of Electroencephalographic Data in Musicians Playing in Ensemble." *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* 47 (9): 1082–90. <https://doi.org/10.1016/j.cortex.2011.05.006>.
- Badino, Leonardo, Alessandro D'Ausilio, Donald Glowinski, Antonio Camurri, and Luciano Fadiga. 2014. "Sensorimotor Communication in Professional Quartets." *Neuropsychologia* 55 (March): 98–104. <https://doi.org/10.1016/j.neuropsychologia.2013.11.012>.
- Bangert, Marc, and Eckart O Altenmuller. 2003. "Mapping Perception to Action in Piano Practice: A Longitudinal DC-EEG Study." *BMC Neuroscience* 4 (October): 26. <https://doi.org/10.1186/1471-2202-4-26>.
- Bardy, B. G., Calabrese, C., Bourgeaud, S., Colomer, C., Pla, S., de Lellis, P., & di Bernardo, M. n.d. "Moving in Unison during and after Visual Interaction." *Submitted*.
- Bardy, Benoît G, Charles P Hoffmann, Bart Moens, Marc Leman, and Simone Dalla Bella. 2015. "Sound-Induced Stabilization of Breathing and Moving." *Annals of the New York Academy of Sciences* 1337 (1): 94–100. <https://doi.org/10.1111/nyas.12650>.
- Barrett, Lisa Feldman, Ralph Adolphs, Stacy Marsella, Aleix M Martinez, and Seth D Pollak. 2019. "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements." *Psychological Science in the Public Interest* 20 (1): 1–68. <https://doi.org/10.1177/1529100619832930>.
- Bartoli, Eleonora, Alessandro D'Ausilio, Jeffrey Berry, Leonardo Badino, Thomas Bever, and Luciano Fadiga. 2015. "Listener-Speaker Perceived Distance Predicts the Degree of Motor Contribution to Speech Perception." *Cerebral Cortex (New York, N.Y. : 1991)* 25 (2): 281–88. <https://doi.org/10.1093/cercor/bht257>.

- Beaudoin, P., Poulin, P., & van de Panne, M. (2007). Adapting wavelet compression to human motion capture clips. In Proceedings of Graphics Interface 2007 (pp. 313-318).
- Becchio, Cristina, Luisa Sartori, Maria Bulgheroni, and Umberto Castiello. 2008. "Both Your Intention and Mine Are Reflected in the Kinematics of My Reach-to-Grasp Movement." *Cognition* 106 (2): 894–912. <https://doi.org/10.1016/j.cognition.2007.05.004>.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798-1828.
- Bernstein, N. 1967. *The Co-Ordination and Regulation of Movements*. Oxford: Pergamon Press.
- Borra, Elena, Marzio Gerbella, Stefano Rozzi, and Giuseppe Luppino. 2017. "The Macaque Lateral Grasping Network: A Neural Substrate for Generating Purposeful Hand Actions." *Neuroscience and Biobehavioral Reviews* 75 (April): 65–90. <https://doi.org/10.1016/j.neubiorev.2017.01.017>.
- Borroni, Paola, Marcella Montagna, Gabriella Cerri, and Fausto Baldissera. 2005. "Cyclic Time Course of Motor Excitability Modulation during the Observation of a Cyclic Hand Movement." *Brain Research* 1065 (1–2): 115–24. <https://doi.org/10.1016/j.brainres.2005.10.034>.
- Bütepage, J., Black, M. J., Kragic, D., & Kjellström, H. (2017). Deep representation learning for human motion prediction and classification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6158-6166).
- Bütepage, J., Ghadirzadeh, A., Karadag, Ö. Ö., Björkman, M., & Kragic, D. (2019). Imitating by generating: deep generative models for imitation of interactive tasks. arXiv preprint arXiv:1910.06031.
- Calvo-Merino, B, D E Glaser, J Grezes, R E Passingham, and P Haggard. 2005. "Action Observation and Acquired Motor Skills: An FMRI Study with Expert Dancers." *Cerebral Cortex (New York, N.Y. : 1991)* 15 (8): 1243–49. <https://doi.org/10.1093/cercor/bhi007>.
- Camponogara, Ivan, Matthew Rodger, Cathy Craig, and Paola Cesari. 2017. "Expert Players Accurately Detect an Opponent's Movement Intentions through Sound Alone." *Journal of Experimental Psychology: Human Perception and Performance* 43 (2): 348–59. <https://doi.org/10.1037/xhp0000316>.
- Candidi, Matteo, Lucia Maria Sachelì, Ilaria Mega, and Salvatore Maria Aglioti. 2014. "Somatotopic Mapping of Piano Fingering Errors in Sensorimotor Experts: TMS Studies in Pianists and Visually Trained Musically Naives." *Cerebral Cortex (New York, N.Y. : 1991)* 24 (2): 435–43. <https://doi.org/10.1093/cercor/bhs325>.
- Cardellicchio, Pasquale, Elisa Dolfini, Pauline M. Hilt, Luciano Fadiga, and Alessandro D'Ausilio. 2020. "Motor Cortical Inhibition during Concurrent Action Execution and Action Observation." *NeuroImage* 208 (116445).
- Casile, Antonino, Eran Dayan, Vittorio Caggiano, Talma Hendler, Tamar Flash, and Martin A Giese. 2010. "Neuronal Encoding of Human Kinematic Invariants during Action Observation." *Cerebral Cortex (New York, N.Y. : 1991)* 20 (7): 1647–55. <https://doi.org/10.1093/cercor/bhp229>.
- Cattaneo, Luigi, Fausto Caruana, Ahmad Jezzini, and Giacomo Rizzolatti. 2009. "Representation of Goal and Movements without Overt Motor Behavior in the Human Motor Cortex: A Transcranial Magnetic Stimulation Study." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 29 (36): 11134–38. <https://doi.org/10.1523/JNEUROSCI.2605-09.2009>.
- Cicchini, Guido Marco, Roberto Arrighi, Luca Cecchetti, Marco Giusti, and David C Burr. 2012. "Optimal Encoding of Interval Timing in Expert Percussionists." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 32 (3): 1056–60. <https://doi.org/10.1523/JNEUROSCI.3411-11.2012>.
- Clark, Andy, and Rick Grush. 1999. "Towards a Cognitive Robotics." *Adaptive Behavior* 7 (1).
- Clayton, Martin, Kelly Jakubowski & Tuomas Eerola, 2019. Interpersonal entrainment in Indian instrumental music performance: Synchronization and movement coordination relate to tempo, dynamics, metrical and cadential structure. *Musicae Scientiae* 23/3, 304–331. [10.1177/1029864919844809](https://doi.org/10.1177/1029864919844809).
- Clayton, Martin, Kelly Jakubowski, Tuomas Eerola, Peter E. Keller, Antonio Camurri, Gualtiero Volpe & Paolo Albornò. submitted. Interpersonal entrainment in music performance: Theory, method and model.
- Codrons, Erwan, Nicolò F Bernardi, Matteo Vandoni, and Luciano Bernardi. 2014. "Spontaneous Group Synchronization of Movements and Respiratory Rhythms." *PLoS ONE* 9 (9). <https://doi.org/10.1371/journal.pone.0107538>.

- Cook, D., Feuz, K. D., & Krishnan, N. C. (2013). Transfer learning for activity recognition: A survey. *Knowledge and information systems*, 36(3), 537-556.
- Cordaro, Daniel T, Rui Sun, Dacher Keltner, Shanmukh Kamble, Niranjan Huddar, and Galen McNeil. 2018. "Universals and Cultural Variations in 22 Emotional Expressions across Five Cultures." *Emotion* 18 (1): 75–93. <https://doi.org/10.1037/emo0000302>.
- Costa, Tommaso, Franco Cauda, Manuella Crini, Mona-Karina Tatu, Alessia Celeghin, Beatrice de Gelder, and Marco Tamietto. 2014. "Temporal and Spatial Neural Dynamics in the Perception of Basic Emotions from Complex Scenes." *Social Cognitive and Affective Neuroscience* 9 (11): 1690–1703. <https://doi.org/10.1093/scan/nst164>.
- Costantini, Marcello, Ettore Ambrosini, and Corrado Sinigaglia. 2012. "Does How I Look at What You're Doing Depend on What I'm Doing?" *Acta Psychologica* 141 (2): 199–204. <https://doi.org/10.1016/j.actpsy.2012.07.012>.
- Coste, Alexandre, Benoit G Bardy, and Ludovic Marin. 2019. "Towards an Embodied Signature of Improvisation Skills." *Frontiers in Psychology* 10: 2441. <https://doi.org/10.3389/fpsyg.2019.02441>.
- Couzin, Iain D, Jens Krause, Nigel R Franks, and Simon A Levin. 2005. "Effective Leadership and Decision-Making in Animal Groups on the Move." *Nature* 433 (7025): 513–16. <https://doi.org/10.1038/nature03236>.
- Cracco, Emiel, Lara Bardi, Charlotte Desmet, Oliver Genschow, Davide Rigoni, Lize De Coster, Ina Radkova, Eliane Deschrijver, and Marcel Brass. 2018. "Automatic Imitation: A Meta-Analysis." *Psychological Bulletin* 144 (5): 453–500. <https://doi.org/10.1037/bul0000143>.
- Cutting, James E, and Lynn T Kozlowski. 1977. "Recognizing Friends by Their Walk: Gait Perception without Familiarity Cues." *Bulletin of the Psychonomic Society* 9 (5): 353–56. <https://doi.org/10.3758/BF03337021>.
- D'Ausilio, A, E Altenmüller, M Olivetti Belardinelli, and M Lotze. 2006. "Cross-Modal Plasticity of the Motor Cortex While Listening to a Rehearsed Musical Piece." *European Journal of Neuroscience* 24 (3): 955–58. <https://doi.org/10.1111/j.1460-9568.2006.04960.x>.
- D'Ausilio, A, L Badino, Y Li, S Tokay, L Craighero, R Canto, Y Aloimonos, and L Fadiga. 2012. "Leadership in Orchestra Emerges from the Causal Relationships of Movement Kinematics." *PLoS ONE* 7 (5).
- D'Ausilio, Alessandro, Leonardo Badino, Yi Li, Sera Tokay, Laila Craighero, Rosario Canto, Yiannis Aloimonos, and Luciano Fadiga. 2012. "Leadership in Orchestra Emerges from the Causal Relationships of Movement Kinematics." Edited by Angela Sirigu. *PLoS ONE* 7 (5): e35757. <https://doi.org/10.1371/journal.pone.0035757>.
- D'Ausilio, Alessandro, Giacomo Novembre, Luciano Fadiga, and Peter E Keller. 2015. "What Can Music Tell Us about Social Interaction?" *Trends in Cognitive Sciences* 19 (3): 111–14. <https://doi.org/10.1016/j.tics.2015.01.005>.
- D'ausilio, Alessandro, Eleonora Bartoli, and Laura Maffongelli. 2015a. "Grasping Synergies: A Motor-Control Approach to the Mirror Neuron Mechanism." *Physics of Life Reviews* 12: 91–103. <https://doi.org/10.1016/j.plrev.2014.11.002>.
- . 2015b. "Motor Control May Support Mirror Neuron Research with New Hypotheses and Methods Reply to Comments on 'Grasping Synergies: A Motor-Control Approach to the Mirror Neuron Mechanism.'" *Physics of Life Reviews* 12: 133–37. <https://doi.org/10.1016/j.plrev.2015.02.005>.
- Dahl, Sofia, and Anders Friberg. 2007. "Visual Perception of Expressiveness in Musicians' Body Movements." *Music Perception* 24 (5): 433–54. <https://doi.org/10.1525/mp.2007.24.5.433>.
- Darwin, Charles. 1872. *The Expression of Emotions in Man and Animals*. London, England: Murray.
- Dayan, Eran, Antonino Casile, Nava Levit-Binnun, Martin A Giese, Talma Hendler, and Tamar Flash. 2007. "Neural Representations of Kinematic Laws of Motion: Evidence for Action-Perception Coupling." *Proceedings of the National Academy of Sciences of the United States of America* 104 (51): 20582–87. <https://doi.org/10.1073/pnas.0710033104>.
- Dikker, Suzanne, Lu Wan, Ido Davidesco, Lisa Kaggen, Matthias Oostrik, James McClintock, Jess Rowland, et al. 2017. "Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom." *Current Biology : CB* 27 (9): 1375–80. <https://doi.org/10.1016/j.cub.2017.04.002>.
- Donnarumma, Francesco, Marcello Costantini, Ettore Ambrosini, Karl Friston, and Giovanni Pezzulo. 2017. "Action Perception as Hypothesis Testing." *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* 89 (April): 45–60. <https://doi.org/10.1016/j.cortex.2017.01.016>.
- Eerola, Tuomas, Kelly Jakubowski, Nikki Moran, Peter E. Keller. & Martin Clayton. (2018). Shared periodic performer

movements coordinate interactions in duo improvisations. *Royal Society Open Science* 5(2), 171520.
<http://dx.doi.org/10.1098/rsos.171520>

- Ekman, Paul, and Wallace V Friesen. 2013. “The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding.” In *Mouton Classics: From Syntax to Cognition. From Phonology to Text*, 819–68. De Gruyter Mouton. <https://doi.org/10.1515/semi.1969.1.1.49>.
- Ekman, Paul, and Dacher Keltner. 1997. “Universal Facial Expressions of Emotion: An Old Controversy and New Findings.” In *Nonverbal Communication: Where Nature Meets Culture.*, 27–46. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Ekman, Paul, Robert W Levenson, and Wallace V Friesen. 1983. “Autonomic Nervous System Activity Distinguishes among Emotions.” *Science* 221 (4616): 1208–10. <https://doi.org/10.1126/science.6612338>.
- Elbert, T, C Pantev, C Wienbruch, B Rockstroh, and E Taub. 1995. “Increased Cortical Representation of the Fingers of the Left Hand in String Players.” *Science (New York, N.Y.)* 270 (5234): 305–7. <https://doi.org/10.1126/science.270.5234.305>.
- Fadiga, L, G Buccino, L Craighero, L Fogassi, V Gallese, and G Pavesi. 1999. “Corticospinal Excitability Is Specifically Modulated by Motor Imagery: A Magnetic Stimulation Study.” *Neuropsychologia* 37 (2): 147–58. [https://doi.org/10.1016/s0028-3932\(98\)00089-x](https://doi.org/10.1016/s0028-3932(98)00089-x).
- Fadiga, L, L Fogassi, G Pavesi, and G Rizzolatti. 1995. “Motor Facilitation during Action Observation: A Magnetic Stimulation Study.” *Journal of Neurophysiology* 73 (6): 2608–11. <https://doi.org/10.1152/jn.1995.73.6.2608>.
- Fairhurst, Merle T, Petr Janata, and Peter E Keller. 2014. “Leading the Follower: An FMRI Investigation of Dynamic Cooperativity and Leader-Follower Strategies in Synchronization with an Adaptive Virtual Partner.” *NeuroImage* 84 (January): 688–97. <https://doi.org/10.1016/j.neuroimage.2013.09.027>.
- Falck-Ytter, Terje, Gustaf Gredeback, and Claes von Hofsten. 2006. “Infants Predict Other People’s Action Goals.” *Nature Neuroscience* 9 (7): 878–79. <https://doi.org/10.1038/nn1729>.
- Finisguerra, Alessandra, Laura Maffongelli, Michela Bassolino, Marco Jacono, Thierry Pozzo, and Alessandro D’Ausilio. 2015. “Generalization of Motor Resonance during the Observation of Hand, Mouth, and Eye Movements.” *Journal of Neurophysiology* 114 (4): 2295–2304. <https://doi.org/10.1152/jn.00433.2015>.
- Flanagan, J Randall, and Roland S Johansson. 2003. “Action Plans Used in Action Observation.” *Nature* 424 (6950): 769–71. <https://doi.org/10.1038/nature01861>.
- Flash, T, and N Hogan. 1985. “The Coordination of Arm Movements: An Experimentally Confirmed Mathematical Model.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 5 (7): 1688–1703.
- Fogassi, Leonardo, Pier Francesco Ferrari, Benno Gesierich, Stefano Rozzi, Fabian Chersi, and Giacomo Rizzolatti. 2005. “Parietal Lobe: From Action Organization to Intention Understanding.” *Science (New York, N.Y.)* 308 (5722): 662–67. <https://doi.org/10.1126/science.1106138>.
- Fox, Nathan A, Marian J Bakermans-Kranenburg, Kathryn H Yoo, Lindsay C Bowman, Erin N Cannon, Ross E Vanderwert, Pier F Ferrari, and Marinus H van IJzendoorn. 2016. “Assessing Human Mirror Activity with EEG Mu Rhythm: A Meta-Analysis.” *Psychological Bulletin* 142 (3): 291–313. <https://doi.org/10.1037/bul0000031>.
- Fragkiadaki, K., Levine, S., Felsen, P., & Malik, J. (2015). Recurrent network models for human dynamics. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4346-4354).
- Frijda, Nico H, and Anna Tcherkassof. 1997. “Facial Expressions as Modes of Action Readiness.” In *The Psychology of Facial Expression*, 78–102. Cambridge University Press. <https://doi.org/10.1017/CBO9780511659911.006>.
- Friston, Karl, Jeremie Mattout, and James Kilner. 2011. “Action Understanding and Active Inference.” *Biological Cybernetics* 104 (1–2): 137–60. <https://doi.org/10.1007/s00422-011-0424-z>.
- Frith, Chris D. 2008. “Social Cognition.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1499): 2033–39. <https://doi.org/10.1098/rstb.2008.0005>.
- Gallese, V, L Fadiga, L Fogassi, and G Rizzolatti. 1996. “Action Recognition in the Premotor Cortex.” *Brain : A Journal of Neurology* 119 (Pt 2 (April): 593–609. <https://doi.org/10.1093/brain/119.2.593>.
- Gangitano, M, F M Mottaghy, and A Pascual-Leone. 2001. “Phase-Specific Modulation of Cortical Motor Output during Movement Observation.” *Neuroreport* 12 (7): 1489–92. <https://doi.org/10.1097/00001756-200105250-00038>.

- Gazzola, Valeria, and Christian Keysers. 2009. "The Observation and Execution of Actions Share Motor and Somatosensory Voxels in All Tested Subjects: Single-Subject Analyses of Unsmoothed fMRI Data." *Cerebral Cortex (New York, N.Y. : 1991)* 19 (6): 1239–55. <https://doi.org/10.1093/cercor/bhn181>.
- Glowinski, Donald, Maurizio Mancini, Roddy Cowie, Antonio Camurri, Carlo Chiorri, and Cian Doherty. 2013. "The Movements Made by Performers in a Skilled Quartet: A Distinctive Pattern, and the Function That It Serves." *Frontiers in Psychology* 4: 841. <https://doi.org/10.3389/fpsyg.2013.00841>.
- Goodfellow, I. and Bengio, Y. and Courville, A. Deep learning. MIT Press. 2016.
- Grafton, Scott T, and Antonia F de C Hamilton. 2007. "Evidence for a Distributed Hierarchy of Action Representation in the Brain." *Human Movement Science* 26 (4): 590–616. <https://doi.org/10.1016/j.humov.2007.05.009>.
- Granger, C. W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37 (3): 424–38.
- Gross, M Melissa, Elizabeth A Crane, and Barbara L Fredrickson. 2010. "Methodology for Assessing Bodily Expression of Emotion." *Journal of Nonverbal Behavior* 34 (4): 223–48. <https://doi.org/10.1007/s10919-010-0094-x>.
- Hallett, Mark. 2007. "Transcranial Magnetic Stimulation: A Primer." *Neuron* 55 (2): 187–99. <https://doi.org/10.1016/j.neuron.2007.06.026>.
- Hardwick, Robert M, Svenja Caspers, Simon B Eickhoff, and Stephan P Swinnen. 2018. "Neural Correlates of Action: Comparing Meta-Analyses of Imagery, Observation, and Execution." *Neuroscience and Biobehavioral Reviews* 94 (November): 31–44. <https://doi.org/10.1016/j.neubiorev.2018.08.003>.
- Hari, R, N Forss, S Avikainen, E Kirveskari, S Salenius, and G Rizzolatti. 1998. "Activation of Human Primary Motor Cortex during Action Observation: A Neuromagnetic Study." *Proceedings of the National Academy of Sciences of the United States of America* 95 (25): 15061–65. <https://doi.org/10.1073/pnas.95.25.15061>.
- Harispe, Sébastien, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain. 2017. "Semantic Similarity from Natural Language and Ontology Analysis," April. <https://doi.org/10.2200/S00639ED1V01Y201504HLT027>.
- Hauelsen, J, and T R Knosche. 2001. "Involuntary Motor Activity in Pianists Evoked by Music Perception." *Journal of Cognitive Neuroscience* 13 (6): 786–92. <https://doi.org/10.1162/08989290152541449>.
- Hilt, P M, B Berret, C Papaxanthis, P J Stapley, and T Pozzo. 2016. "Evidence for Subjective Values Guiding Posture and Movement Coordination in a Free-Endpoint Whole-Body Reaching Task." *Scientific Reports* 6 (April): 23868. <https://doi.org/10.1038/srep23868>.
- Hilt, Pauline M., Pasquale Cardellicchio, Elisa Dolfini, Thierry Pozzo, Luciano Fadiga, and Alessandro D'Ausilio. n.d. "Motor Recruitment during Action Observation: Effect of Interindividual Differences in Action Strategy." *Cerebral Cortex (New York, N.Y. : 1991)*2.
- Hilt, Pauline M, Leonardo Badino, Alessandro D'Ausilio, Gualtiero Volpe, Sera Tokay, Luciano Fadiga, and Antonio Camurri. 2019. "Multi-Layer Adaptation of Group Coordination in Musical Ensembles." *Scientific Reports* 9 (1): 5854. <https://doi.org/10.1038/s41598-019-42395-4>.
- Hilt, Pauline M, Eleonora Bartoli, Elisabetta Ferrari, Marco Jacono, Luciano Fadiga, and Alessandro D'Ausilio. 2017. "Action Observation Effects Reflect the Modular Organization of the Human Motor System." *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* 95 (October): 104–18. <https://doi.org/10.1016/j.cortex.2017.07.020>.
- Homma, Ikuo, and Yuri Masaoka. 2008. "Breathing Rhythms and Emotions." *Experimental Physiology* 93 (9): 1011–21. <https://doi.org/10.1113/expphysiol.2008.042424>.
- Horak, F B, and L M Nashner. 1986. "Central Programming of Postural Movements: Adaptation to Altered Support-Surface Configurations." *Journal of Neurophysiology* 55 (6): 1369–81. <https://doi.org/10.1152/jn.1986.55.6.1369>.
- Hoshiyama, M, Y Kitamura, S Koyama, S Watanabe, M Shimojo, and R Kakigi. 1996. "Reciprocal Change of Motor Evoked Potentials Preceding Voluntary Movement in Humans." *Muscle & Nerve* 19 (2): 125–31. [https://doi.org/10.1002/\(SICI\)1097-4598\(199602\)19:2<125::AID-MUS1>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-4598(199602)19:2<125::AID-MUS1>3.0.CO;2-G).
- Hurley, Susan. 2008. "The Shared Circuits Model (SCM): How Control, Mirroring, and Simulation Can Enable Imitation, Deliberation, and Mindreading." *The Behavioral and Brain Sciences* 31 (1): 1–58. <https://doi.org/10.1017/S0140525X07003123>.

- Iacoboni, M, R P Woods, M Brass, H Bekkering, J C Mazziotta, and G Rizzolatti. 1999. "Cortical Mechanisms of Human Imitation." *Science (New York, N.Y.)* 286 (5449): 2526–28. <https://doi.org/10.1126/science.286.5449.2526>.
- Iacoboni, Marco, Istvan Molnar-Szakacs, Vittorio Gallese, Giovanni Buccino, John C Mazziotta, and Giacomo Rizzolatti. 2005. "Grasping the Intentions of Others with One's Own Mirror Neuron System." *PLoS Biology* 3 (3): e79. <https://doi.org/10.1371/journal.pbio.0030079>.
- Insafutdinov, Eldar, Leonid Pishchulin, Bjoern Andres, Mykhaylo Andriluka, and Bernt Schiele. 2016. "DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model BT - Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI." *Arxiv*, 34–50. https://doi.org/10.1007/978-3-319-46466-4_3.
- Izard, Carroll E (Carroll Ellis). 1971. *The Face of Emotion*. Appleton-Century-Crofts.
- Jackendoff, Ray. 2009. *Language, Consciousness, Culture : Essays on Mental Structure*. MIT Press.
- Janssen, Daniel, Wolfgang I Schöllhorn, Jessica Lubienetzki, Karina Fölling, Henrike Kokenge, and Keith Davids. 2008. "Recognition of Emotions in Gait Patterns by Means of Artificial Neural Nets." *Journal of Nonverbal Behavior* 32 (2): 79–92. <https://doi.org/10.1007/s10919-007-0045-3>.
- Jiu, Mingyuan, Nelly Pustelnik, Stefan Janaqi, Mériam Chebre, Lin Qi, and Philippe Ricoux. 2019. "Sparse Hierarchical Interaction Learning with Epigraphical Projection." *Journal of Signal Processing Systems*, November. <https://doi.org/10.1007/s11265-019-01478-1>.
- Johansson, Gunnar. 1973. "Visual Perception of Biological Motion and a Model for Its Analysis." *Perception & Psychophysics* 14 (2): 201–11. <https://doi.org/10.3758/BF03212378>.
- Jyoti Joshi, Abhinav Dhall, Roland Goecke, and Jeffrey F. Cohn. 2013. Relative body parts movement for automatic depression analysis. In *Proceedings of ACII*, 492-497.
- Kaplan, Harris S, Oriana Salazar Thula, Niklas Khoss, and Manuel Zimmer. 2019. "Nested Neuronal Dynamics Orchestrate a Behavioral Hierarchy across Timescales (Review)." *Neuron*, 1–15. <https://doi.org/10.1016/j.neuron.2019.10.037>.
- Kilner, James M, Alice Neal, Nikolaus Weiskopf, Karl J Friston, and Chris D Frith. 2009. "Evidence of Mirror Neurons in Human Inferior Frontal Gyrus." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 29 (32): 10153–59. <https://doi.org/10.1523/JNEUROSCI.2668-09.2009>.
- Kirschner, Sebastian, and Tomasello. 2010. "Joint Music Making Promotes Prosocial Behavior in 4-Year-Old Children." *Evol. Hum. Behav.* 31: 354–364.
- Kleinsmith, Andrea, and Nadia Bianchi-Berthouze. 2013. "Affective Body Expression Perception and Recognition: A Survey." *IEEE Transactions on Affective Computing* 4 (1): 15–33. <https://doi.org/10.1109/T-AFFC.2012.16>.
- Kingma, D. P. & Welling, M. (2014). Auto-Encoding Variational Bayes. In *Proceedings of International Conference on Learning Representations (ICLR)*.
- Knoblich, Gunther, and Jerome Scott Jordan. 2003. "Action Coordination in Groups and Individuals: Learning Anticipatory Control." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 29 (5): 1006–16. <https://doi.org/10.1037/0278-7393.29.5.1006>.
- Kokal, Idil, Annerose Engel, Sebastian Kirschner, and Christian Keysers. 2011. "Synchronized Drumming Enhances Activity in the Caudate and Facilitates Prosocial Commitment--If the Rhythm Comes Easily." *PLoS One* 6 (11): e27272. <https://doi.org/10.1371/journal.pone.0027272>.
- Kokal, Idil, and Christian Keysers. 2010. "Granger Causality Mapping during Joint Actions Reveals Evidence for Forward Models That Could Overcome Sensory-Motor Delays." *PLoS One* 5 (10): e13507. <https://doi.org/10.1371/journal.pone.0013507>.
- Komeilipoor, Naeem, Fabio Pizzolato, Andreas Daffertshofer, and Paola Cesari. 2013. "Excitability of Motor Cortices as a Function of Emotional Sounds." *PLoS ONE* 8 (5). <https://doi.org/10.1371/journal.pone.0063060>.
- Kong, Y., Tao, Z., & Fu, Y. (2020). Adversarial action prediction networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(3), 539-553.
- Kourtis, Dimitrios, Natalie Sebanz, and Gunther Knoblich. 2010. "Favouritism in the Motor System: Social Interaction Modulates Action Simulation." *Biology Letters* 6 (6): 758–61. <https://doi.org/10.1098/rsbl.2010.0478>.

- Kuppens, Peter, Francis Tuerlinckx, James A Russell, and Lisa Feldman Barrett. 2013. "The Relation between Valence and Arousal in Subjective Experience." *Psychological Bulletin* 139 (4): 917–40. <https://doi.org/10.1037/a0030811>.
- la Rosa, Stephan de, Ylva Ferstl, and Heinrich H Bulthoff. 2016. "Visual Adaptation Dominates Bimodal Visual-Motor Action Adaptation." *Scientific Reports* 6 (March): 23829. <https://doi.org/10.1038/srep23829>.
- Li, K., & Fu, Y. (2014). Prediction of human activity by discovering temporal sequence patterns. *IEEE transactions on pattern analysis and machine intelligence*, 36(8), 1644-1657.
- Li, Y., Fermüller, C., Aloimonos, Y., & Ji, H. (2010). Learning shift-invariant sparse representation of actions. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 2630-2637).
- Limb, Charles J, and Allen R Braun. 2008. "Neural Substrates of Spontaneous Musical Performance: An FMRI Study of Jazz Improvisation." *PLoS One* 3 (2): e1679. <https://doi.org/10.1371/journal.pone.0001679>.
- Liu, J., Wang, G., Duan, L. Y., Abdiyeva, K., & Kot, A. C. (2017). Skeleton-based human action recognition with global context-aware attention LSTM networks. *IEEE Transactions on Image Processing*, 27(4), 1586-1599.
- Loehr, Janeen D, Dimitrios Kourtis, Cordula Vesper, Natalie Sebanz, and Gunther Knoblich. 2013. "Monitoring Individual and Joint Action Outcomes in Duet Music Performance." *Journal of Cognitive Neuroscience* 25 (7): 1049–61. https://doi.org/10.1162/jocn_a_00388.
- Lotze, M, G Scheler, H-R M Tan, C Braun, and N Birbaumer. 2003. "The Musician's Brain: Functional Imaging of Amateurs and Professionals during Performance and Imagery." *NeuroImage* 20 (3): 1817–29. <https://doi.org/10.1016/j.neuroimage.2003.07.018>.
- Loula, Fani, Sapna Prasad, Kent Harber, and Maggie Shiffrar. 2005. "Recognizing People from Their Movement." *Journal of Experimental Psychology: Human Perception and Performance* 31 (1): 210–20. <https://doi.org/10.1037/0096-1523.31.1.210>.
- Lu, C., & Ferrier, N. J. (2004). Repetitive motion analysis: Segmentation and event classification. *IEEE transactions on pattern analysis and machine intelligence*, 26(2), 258-263.
- Lucy, P., Cohn, J. F., Prkachin, K. M., Solomon, P., & Matthews, I. (2011). Painful data: The UNBC-McMaster Shoulder Pain Expression Archive Database. *IEEE International Conference on Automatic Face and Gesture Recognition (FG2011)*.
- Lussu, Vincenzo, Radoslaw Niewiadomski, Gualtiero Volpe, and Antonio Camurri. 2019. "The Role of Respiration Audio in Multimodal Analysis of Movement Qualities." *Journal on Multimodal User Interfaces*.
- Mather, G, and L Murdoch. 1994. "Gender Discrimination in Biological Motion Displays Based on Dynamic Cues." *Proceedings of the Royal Society B: Biological Sciences* 258 (1353): 273–79. <https://doi.org/10.1098/rspb.1994.0173>.
- Matsumoto, David, and Bob Willingham. 2009. "Spontaneous Facial Expressions of Emotion of Congenitally and Noncongenitally Blind Individuals." *Journal of Personality and Social Psychology* 96 (1): 1–10. <https://doi.org/10.1037/a0014037>.
- Meijer, Marco de. 1989. "The Contribution of General Features of Body Movement to the Attribution of Emotions." *Journal of Nonverbal Behavior* 13 (4): 247–68. <https://doi.org/10.1007/BF00990296>.
- Mehta, D., Rhodin, H., Casas, D., Fua, P., Sotnychenko, O., Xu, W., & Theobalt, C. (2017). Monocular 3d human pose estimation in the wild using improved CNN supervision. In 2017 International Conference on 3D Vision (3DV) (pp. 506-516).
- Molnar, C. Interpretable machine learning. Lulu. com. 2019.
- Mukamel, Roy, Arne D Ekstrom, Jonas Kaplan, Marco Iacoboni, and Itzhak Fried. 2010. "Single-Neuron Responses in Humans during Execution and Observation of Actions." *Current Biology : CB* 20 (8): 750–56. <https://doi.org/10.1016/j.cub.2010.02.045>.
- Murata, A, L Fadiga, L Fogassi, V Gallese, V Raos, and G Rizzolatti. 1997. "Object Representation in the Ventral Premotor Cortex (Area F5) of the Monkey." *Journal of Neurophysiology* 78 (4): 2226–30. <https://doi.org/10.1152/jn.1997.78.4.2226>.
- Nagy, Mate, Zsuzsa Akos, Dora Biro, and Tamas Vicsek. 2010. "Hierarchical Group Dynamics in Pigeon Flocks." *Nature* 464 (7290): 890–93. <https://doi.org/10.1038/nature08891>.

- Naish, Katherine R, Carmel Houston-Price, Andrew J Bremner, and Nicholas P Holmes. 2014. "Effects of Action Observation on Corticospinal Excitability: Muscle Specificity, Direction, and Timing of the Mirror Response." *Neuropsychologia* 64 (November): 331–48. <https://doi.org/10.1016/j.neuropsychologia.2014.09.034>.
- Nath, Tanmay, Alexander Mathis, An Chi Chen, Amir Patel, Matthias Bethge, and Mackenzie Weygandt Mathis. 2018. "Using DeepLabCut for 3D Markerless Pose Estimation across Species and Behaviors." *BioRxiv*, 476531. <https://doi.org/10.1101/476531>.
- Newbold, J., Bianchi-Berthouze, N., Gold, N. E., Tajadura-Jimenez, A., & Williams, A. (2016). Musically Informed Sonification for Chronic Pain Rehabilitation: Facilitating Progress & Avoiding Over-doing. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 5685-5697. doi:10.1145/2858036.2858302
- Newman-Norlund, Roger D, Hein T van Schie, Alexander M J van Zuijlen, and Harold Bekkering. 2007. "The Mirror Neuron System Is More Active during Complementary Compared with Imitative Action." *Nature Neuroscience* 10 (7): 817–18. <https://doi.org/10.1038/nn1911>.
- Novembre, Giacomo, Luca F Ticini, Simone Schutz-Bosbach, and Peter E Keller. 2012. "Distinguishing Self and Other in Joint Action. Evidence from a Musical Paradigm." *Cerebral Cortex (New York, N.Y. : 1991)* 22 (12): 2894–2903. <https://doi.org/10.1093/cercor/bhr364>.
- . 2014. "Motor Simulation and the Coordination of Self and Other in Real-Time Joint Action." *Social Cognitive and Affective Neuroscience* 9 (8): 1062–68. <https://doi.org/10.1093/scan/nst086>.
- T. A. Olugbade, N. Bianchi-Berthouze, N. Marquardt, and A. C. Williams, "Human Observer and Automatic Assessment of Movement Related Self-Efficacy in Chronic Pain: from Exercise to Functional Activity," IEEE Transact Affective Comput (Early Access), 2018, doi: 10.1109/TAFFC.2018.2798576.
- Oneto, L. Model Selection and Error Estimation in a Nutshell. Springer. 2020.
- Pantev, C, R Oostenveld, A Engelen, B Ross, L E Roberts, and M Hoke. 1998. "Increased Auditory Cortical Representation in Musicians." *Nature* 392 (6678): 811–14. <https://doi.org/10.1038/33918>.
- Pecenka, Nadine, and Peter E Keller. 2011. "The Role of Temporal Prediction Abilities in Interpersonal Sensorimotor Synchronization." *Experimental Brain Research* 211 (3–4): 505–15. <https://doi.org/10.1007/s00221-011-2616-0>.
- Petrides, Michael, Genevieve Cadoret, and Scott Mackey. 2005. "Orofacial Somatomotor Responses in the Macaque Monkey Homologue of Broca's Area." *Nature* 435 (7046): 1235–38. <https://doi.org/10.1038/nature03628>.
- Petrides, Michael, and Deepak N Pandya. 2009. "Distinct Parietal and Temporal Pathways to the Homologues of Broca's Area in the Monkey." *PLoS Biology* 7 (8): e1000170. <https://doi.org/10.1371/journal.pbio.1000170>.
- Pezzulo, Giovanni, Francesco Donnarumma, Haris Dindo, Alessandro D'Ausilio, Ivana Konvalinka, and Cristiano Castelfranchi. 2019. "The Body Talks: Sensorimotor Communication and Its Brain and Kinematic Signatures." *Physics of Life Reviews* 28 (March): 1–21. <https://doi.org/10.1016/j.plrev.2018.06.014>.
- Plutchik, R. 1962. *The Emotions: Facts Theories and a New Model*. New York: Random House.
- Pollick, F E, H M Paterson, A Bruderlin, and A J Sanford. 2001. "Perceiving Affect from Arm Movement." *Cognition* 82 (2): B51–61. [https://doi.org/10.1016/S0010-0277\(01\)00147-0](https://doi.org/10.1016/S0010-0277(01)00147-0).
- Prabhu, Gita, Martin Voss, Thomas Brochier, Luigi Cattaneo, Patrick Haggard, and Roger Lemon. 2007. "Excitability of Human Motor Cortex Inputs Prior to Grasp." *The Journal of Physiology* 581 (Pt 1): 189–201. <https://doi.org/10.1113/jphysiol.2006.123356>.
- Ragert, Marie, Tim Schroeder, and Peter E Keller. 2013. "Knowing Too Little or Too Much: The Effects of Familiarity with a Co-Performer's Part on Interpersonal Coordination in Musical Ensembles." *Frontiers in Psychology* 4: 368. <https://doi.org/10.3389/fpsyg.2013.00368>.
- Rands, Sean A, Guy Cowlshaw, Richard A Pettifor, J Marcus Rowcliffe, and Rufus A Johnstone. 2003. "Spontaneous Emergence of Leaders and Followers in Foraging Pairs." *Nature* 423 (6938): 432–34. <https://doi.org/10.1038/nature01630>.
- Richardson, Michael J, Kerry L Marsh, Robert W Isenhower, Justin R L Goodman, and R C Schmidt. 2007. "Rocking Together: Dynamics of Intentional and Unintentional Interpersonal Coordination." *Human Movement Science* 26 (6): 867–91. <https://doi.org/10.1016/j.humov.2007.07.002>.
- Rivas, J. J., Orihuela-Espina, F., Palafox, L., Berthouze, N. L., del Carmen Lara, M., Herdandez-Franco, J., & Sucar, E.

- (2018). Unobtrusive Inference of Affective States in Virtual Rehabilitation from Upper Limb Motions: A Feasibility Study. *IEEE Transactions on Affective Computing*.
- Rizzolatti, G, L Fadiga, M Matelli, V Bettinardi, E Paulesu, D Perani, and F Fazio. 1996. "Localization of Grasp Representations in Humans by PET: 1. Observation versus Execution." *Experimental Brain Research* 111 (2): 246–52. <https://doi.org/10.1007/bf00227301>.
- Rizzolatti, Giacomo, and Corrado Sinigaglia. 2016. "The Mirror Mechanism: A Basic Principle of Brain Function." *Nature Reviews. Neuroscience* 17 (12): 757–65. <https://doi.org/10.1038/nrn.2016.135>.
- Roether, Claire L, Lars Omlor, Andrea Christensen, and Martin A Giese. 2009. "Critical Features for the Perception of Emotion from Gait." *Journal of Vision* 9 (6): 1–32. <https://doi.org/10.1167/9.6.15>.
- Rothwell, J C. 1997. "Techniques and Mechanisms of Action of Transcranial Stimulation of the Human Motor Cortex." *Journal of Neuroscience Methods* 74 (2): 113–22. [https://doi.org/10.1016/s0165-0270\(97\)02242-5](https://doi.org/10.1016/s0165-0270(97)02242-5).
- Rounis, E, V van Polanen, and M Davare. 2018. "A Direct Effect of Perception on Action When Grasping a Cup." *Scientific Reports* 8 (1): 171. <https://doi.org/10.1038/s41598-017-18591-5>.
- Russell, J A, and J M F Dols. 1997. *The Psychology of Facial Expression*. <http://www.ffri.hr/%7B~%7Dibrdar/komunikacija/seminari/Russell,>
- Ryoo, M. S. (2011). Human activity prediction: Early recognition of ongoing activities from streaming videos. In 2011 International Conference on Computer Vision (pp. 1036-1043).
- Sacharin, V, K Schlegel, and K R Scherer. 2012. "Geneva Emotion Wheel Rating Study." <http://www.affective-sciences.org/researchmaterial>.
- Sanger, Johanna, Viktor Muller, and Ulman Lindenberger. 2012. "Intra- and Interbrain Synchronization and Network Properties When Playing Guitar in Duets." *Frontiers in Human Neuroscience* 6: 312. <https://doi.org/10.3389/fnhum.2012.00312>.
- . 2013. "Directionality in Hyperbrain Networks Discriminates between Leaders and Followers in Guitar Duets." *Frontiers in Human Neuroscience* 7: 234. <https://doi.org/10.3389/fnhum.2013.00234>.
- Sawada, Misako, Kazuhiro Suda, and Motonobu Ishii. 2003. "Expression of Emotions in Dance: Relation between Arm Movement Characteristics and Emotion." *Perceptual and Motor Skills* 97 (3 I): 697–708. <https://doi.org/10.2466/pms.2003.97.3.697>.
- Schaal, S. (2006). Dynamic movement primitives—a framework for motor control in humans and humanoid robotics. In *Adaptive motion of animals and machines* (pp. 261-280).
- Schilbach, Leonhard, Bert Timmermans, Vasudevi Reddy, Alan Costall, Gary Bente, Tobias Schlicht, and Kai Vogeley. 2013. "Toward a Second-Person Neuroscience." *The Behavioral and Brain Sciences* 36 (4): 393–414. <https://doi.org/10.1017/S0140525X12000660>.
- Schlaug, G, L Jancke, Y Huang, and H Steinmetz. 1995. "In Vivo Evidence of Structural Brain Asymmetry in Musicians." *Science (New York, N.Y.)* 267 (5198): 699–701. <https://doi.org/10.1126/science.7839149>.
- Schmitz, Judith, Eleonora Bartoli, Laura Maffongelli, Luciano Fadiga, Nuria Sebastian-Galles, and Alessandro D'Ausilio. 2019. "Motor Cortex Compensates for Lack of Sensory and Motor Experience during Auditory Speech Perception." *Neuropsychologia* 128 (May): 290–96. <https://doi.org/10.1016/j.neuropsychologia.2018.01.006>.
- Sebanz, Natalie, Harold Bekkering, and Gunther Knoblich. 2006. "Joint Action: Bodies and Minds Moving Together." *Trends in Cognitive Sciences* 10 (2): 70–76. <https://doi.org/10.1016/j.tics.2005.12.009>.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).
- Shalev-Shwartz, S. and Ben-David, S. *Understanding machine learning: From theory to algorithms*. Cambridge University Press. 2014.
- Silvanto, Juha, Neil Muggleton, and Vincent Walsh. 2008. "State-Dependency in Brain Stimulation Studies of Perception and Cognition." *Trends in Cognitive Sciences* 12 (12): 447–54. <https://doi.org/10.1016/j.tics.2008.09.004>.
- Slowinski, Piotr, Chao Zhai, Francesco Alderisio, Robin Salesse, Mathieu Gueugnon, Ludovic Marin, Benoit G Bardy,

- Mario di Bernardo, and Krasimira Tsaneva-Atanasova. 2016. “Dynamic Similarity Promotes Interpersonal Coordination in Joint Action.” *Journal of the Royal Society, Interface* 13 (116). <https://doi.org/10.1098/rsif.2015.1093>.
- Słowiński, Piotr, Chao Zhai, Francesco Alderisio, Robin Salesse, Mathieu Gueugnon, Ludovic Marin, Benoit G Bardy, Mario Di Bernardo, and Krasimira Tsaneva-Atanasova. 2016. “Dynamic Similarity Promotes Interpersonal Coordination in Joint Action.” *Journal of the Royal Society Interface* 13 (116). <https://doi.org/10.1098/rsif.2015.1093>.
- Soriano, Marco, Andrea Cavallo, Alessandro D’Ausilio, Cristina Becchio, and Luciano Fadiga. 2018. “Movement Kinematics Drive Chain Selection toward Intention Detection.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (41): 10452–57. <https://doi.org/10.1073/pnas.1809825115>.
- Stergiou, A., Kapidis, G., Kalliatakis, G., Chrysoulas, C., Veltkamp, R., & Poppe, R. (2019). Saliency tubes: Visual explanations for spatio-temporal convolutions. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 1830-1834).
- Tami, Myriam, Marianne Clausel, Emilie Devijver, Adrien Dulac, Eric Gaussier, Stefan Janaqi, and Meriam Chebre. 2018. “Uncertain Trees: Dealing with Uncertain Inputs in Regression Trees,” October. <http://arxiv.org/abs/1810.11698>.
- Troje, Nikolaus F, and Cord Westhoff. 2006. “The Inversion Effect in Biological Motion Perception: Evidence for a ‘Life Detector’?” *Current Biology* 16 (8): 821–24. <https://doi.org/10.1016/j.cub.2006.03.022>.
- Uhlig, Marie, Merle T Fairhurst, and Peter E Keller. 2013. “The Importance of Integration and Top-down Saliency When Listening to Complex Multi-Part Musical Stimuli.” *NeuroImage* 77 (August): 52–61. <https://doi.org/10.1016/j.neuroimage.2013.03.051>.
- Ulzen, Niek R van, Claudine J C Lamoth, Andreas Daffertshofer, Gun R Semin, and Peter J Beek. 2008. “Characteristics of Instructed and Uninstructed Interpersonal Coordination While Walking Side-by-Side.” *Neuroscience Letters* 432 (2): 88–93. <https://doi.org/10.1016/j.neulet.2007.11.070>.
- Verduyn, Philippe, Pauline Delaveau, Jean-Yves Rotgé, Philippe Fossati, and Iven Van Mechelen. 2015. “Determinants of Emotion Duration and Underlying Psychological and Neural Mechanisms.” *Emotion Review* 7 (4): 330–35. <https://doi.org/10.1177/1754073915590618>.
- Vergotte, Grégoire, Stéphane Perrey, Muthuraman Muthuraman, Stefan Janaqi, and Kjerstin Torre. 2018. “Concurrent Changes of Brain Functional Connectivity and Motor Variability When Adapting to Task Constraints.” *Frontiers in Physiology* 9 (July). <https://doi.org/10.3389/fphys.2018.00909>.
- Wing, Alan M, Satoshi Endo, Adrian Bradbury, and Dirk Vorberg. 2014. “Optimal Feedback Correction in String Quartet Synchronization.” *Journal of the Royal Society, Interface* 11 (93): 20131125. <https://doi.org/10.1098/rsif.2013.1125>.
- Wolpert, D M, Z Ghahramani, and M I Jordan. 1995. “An Internal Model for Sensorimotor Integration.” *Science (New York, N.Y.)* 269 (5232): 1880–82. <https://doi.org/10.1126/science.7569931>.
- Wolpert, D M, and M Kawato. 1998. “Multiple Paired Forward and Inverse Models for Motor Control.” *Neural Networks : The Official Journal of the International Neural Network Society* 11 (7–8): 1317–29. [https://doi.org/10.1016/s0893-6080\(98\)00066-5](https://doi.org/10.1016/s0893-6080(98)00066-5).
- Wolpert, Daniel M, Kenji Doya, and Mitsuo Kawato. 2003. “A Unifying Computational Framework for Motor Control and Social Interaction.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 358 (1431): 593–602. <https://doi.org/10.1098/rstb.2002.1238>.
- Young, William, Matthew Rodger, and Cathy M Craig. 2013. “Perceiving and Reenacting Spatiotemporal Characteristics of Walking Sounds.” *Journal of Experimental Psychology: Human Perception and Performance* 39 (2): 464–76. <https://doi.org/10.1037/a0029402>.
- Zatorre, Robert J, Joyce L Chen, and Virginia B Penhune. 2007. “When the Brain Plays Music: Auditory-Motor Interactions in Music Perception and Production.” *Nature Reviews. Neuroscience* 8 (7): 547–58. <https://doi.org/10.1038/nrn2152>.
- Zhou, X., Huang, Q., Sun, X., Xue, X., & Wei, Y. (2017). Towards 3d human pose estimation in the wild: a weakly-supervised approach. In Proceedings of the IEEE International Conference on Computer Vision (pp. 398-407).