# D2.2

# Results on prediction in action execution and observation – Phase 2

| | |
|---|---|
| Project No | GA824160 |
| Project Acronym | EnTimeMent |
| Project full title | ENtrainment & synchronization at multiple TIME scales in the MENTal foundations of expressive gesture |
| Instrument | FET Proactive |
| Type of action | RIA |
| Start Date of project | 1 January 2019 |
| Duration | 48 months |
| WP responsible | IIT |
| Due date | Month 30 |

# Table of Contents

# Introduction

This deliverable reports on the progress on the research conducted between M7-M18 of the EnTimeMent project with regards to the individual action execution and observation axis, focused on studies on individual motor behaviour. This part of the project constitute the baseline for research and theoretical work developed in the Phase I of the EnTimeMent project, focused on dyadic studies (n=2) and on group motor behaviour (involving three people or more, n>2) which are reported in D2.3 and D2.5 respectively. The numbering of the studies reported herein, refers to the most recent version of deliverable D1.2 Research Requirements providing an update on the methodological background and know-how of the studies. In this deliverable we report results of studies that have finished the stage of data collection and analysis (*2.1.2, 2.1.6, 2.1.8, 2.1.11, 2.1.15*).

A major theoretical shift in cognitive neuroscience was driven by a new conceptualization of the motor system. In fact, motor processes seem to 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. Critically, the recruitment of motor programs, during action/object perception, constrain the active search of specific sensory features that maximize the discrimination between different perceptual hypotheses and support prediction of future information at multiple timescales. 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.

In this deliverable we collected all ongoing research, investigating action-perception coupling in single individuals and thus on the neurobehavioral building blocks allowing sensorimotor communication in dyads or groups. Studies presented below are those that have either published or are in an advanced stage close to submission for publication.

The first two studies are based on the same theoretical framework suggesting that Individual Motor Signatures (IMS) characterize action execution. Briefly, IMSs are relatively stable movement strategies that each one of us unknowingly display when moving in our environment. Interestingly, data from the first experiment (*2.1.2*) provide evidence that motor activations during action observation are driven by the mismatch between the observers' and actors' IMSs. The larger the distance the larger is the motor recruitment, thus suggesting that the motor system might indeed act as an inferential engine that compares other's action to our own template. The second study (*2.1.6*) approaches a similar problem from a different perspective. In fact, the goal of this project is to automatically extract IMSs from arm movement by using both traditional machine learning methods and more recent deep learning strategies.

The third *(2.1.8)* and fourth projects *(2.1.11)* explored the sensorimotor bases of expressivity. Among them the first aimed at extracting expressivity measures from complex individual body motion. In order to do so, a novel computational pipeline has been implemented by integrating graph and game theory towards the analysis of the perceived origin of full-body human movement and its propagation. In fact, the analysis of the origin of movement is an important component in the understanding and modeling expressivity. The data, extracted with the computational method, have then been submitted for evaluation to a panel of dancers with varying degrees of expertise. The following study has instead tried to discover which specific postural and kinematic features could be computed from affective whole-body movement videos and related those to brain responses. By means of state-of-the-art neuroimaging methods it was investigated whether the (dis)similarity of body posture and kinematics between different emotional categories could explain neural responses to body expressions in and beyond body-selective regions.

Finally, the last section reports on the *2.1.14* research activities aimed at the automatic detection of pain and associated behavior from body movements. This research program is a key component of WP4, constituting one of the use case scenarios planned in

EnTimeMent. Taken together, Phase I results reported herein pushed forward current state-of-the-art description of human movement from the perspective of individual differences (IMS) and their expressive properties. Studies reported below addressed multiple gaps in the body of research and emphasized the importance of approaching human movement analysis and modeling through the lens of mid-layer features. Research roadmap for Phase II of the EnTimeMent project has been established (D1.2 Research requirements), which will push further the frontiers towards a full understanding of the importance of modeling human movement across multiple timescales.

## Updates from Phase 1 to Phase 2

UCL - Added a fourth study titled 'Accounting for Timescale Differences in Leveraging Human Activity Recognition (HAR) to enable Protective Behaviour Detection (PBD) with An Hierarchical HAR-PBD Model'
UNIGE – Added new descriptions and results in the section "2.1.8 Perception of the origin of full body human movement and its propagation" and section "2.1.6 Investigate singularity in ellipses drawing".
IIT-UNIFE – Added results from one project as described in Deliverable 1.2, point 2.1.1 "Cortico-motor alpha coherence influence visual perception".
IIT GE – Added results from a study aimed at investigating how action and perception intersect at the single-trial level in autism spectrum disorders

## 2.1.2 Action variability in action observation and execution

**For a full description please see:** Hilt P. M., Cardellicchio P., Dolfini E., Pozzo T., Fadiga L., D'Ausilio A. (2020) Motor recruitment during action observation: effect of interindividual differences in action strategy. Cereb Cortex, 30(7), 3910–3920.

### Action Perception

Mirror neurons were originally described as visuomotor neurons that are engaged both during visual presentation of actions performed by conspecifics, and during the actual execution of these actions (Rizzolatti and Craighero 2004). These neurons were first discovered using single-cell recordings in monkey premotor cortex (area F5; di Pellegrino et al. 1992) and later within monkey inferior parietal cortex (PF/PFG; Gallese et al. 2002; Fogassi et al. 2005).

Since then, there has been a growing interest in mirror neurons both in the scientific literature and the popular media. The widespread interest was in particular driven by their potential role in imitation and thus in a fundamental aspect of social cognition (Iacoboni 2005; Rizzolatti and Sinigaglia 2010). In follow-up studies, neurons with mirror properties have been found in different parietal and frontal areas of monkeys and other species, including humans (Rizzolatti and Sinigaglia 2016).

The mirror neuron system has also been associated with action perception. In fact, others' action anticipation and comprehension might be achieved both by the ventral route (Middle Temporal Gyrus – MTG - and the anterior Inferior Frontal Gyrus - aIFG), and the dorsal route (Inferior Parietal Lobule – IPL - and the posterior Inferior Frontal Gyrus - pIFG). The dorsal stream may support this process by reactivating the most likely action needed to achieve the predicted goal. In line with this account, action discrimination could rely on internal forward models (Flanagan and Johansson 2003; Kilner et al. 2004) to anticipate the unfolding of a given action (Schütz-Bosbach and Prinz 2007).

## Mirror neuron system in humans

Immediately following the initial reports of mirror neurons in the macaque brain, the existence of an analogous mechanism in humans was discussed. While some authors argued that clear evidence of a human mirror neuron system was still lacking (e.g. Dinstein 2008; Lingnau et al. 2009; Turella et al. 2009), further and numerous results coming from various techniques such as transcranial magnetic stimulation (TMS; Fadiga et al. 2005; 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 (Rizzolatti and Sinigaglia 2010).

Based on human brain-imaging data (Rizzolatti et al. 1996; Decety et al. 1997; Iacoboni et al. 1999) and cytoarchitecture (Petrides 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 to the monkey mirror area PF/PFG (Rizzolatti et al. 2001; Rizzolatti and Craighero 2004).

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 (Pineda 2008; Fox et al. 2016). These results suggest that Rolandic mu event-related desynchronization (Cochin et al. 1998; Babiloni et al. 2002) during action observation reflects activity of a mirror-like system present in humans (Sebastiani et al. 2014; Fox et al. 2016; Lapenta et al. 2018).

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 (Fadiga et al. 1995; for a review please see: Fadiga et al. 2005; Naish et al. 2014; D’Ausilio et al. 2015).

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; Blakemore and Frith 2005; Falck-Ytter et al. 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 Hogans 1985; Bennequin et al. 2009) could support inferences about the unfolding of other's actions (Dayan et al. 2007; Casile et al. 2010). 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; Rizzolatti et al. 2001; Rizzolatti and Craighero 2004; Rizzolatti and Sinigaglia 2016). 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 (Kilner 2012; D'Ausilio et al. 2015; 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 (Rizzolatti et al. 2001; Rizzolatti and Craighero 2004; 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 amounts to finding a solution to a many-to-many mapping problem.

Here we suggest that a simpler mapping exists between behavioral goals and the lower dimensionality space of whole-body configurations (i.e. synergies; 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; Berret et al. 2009). On the top of that, each individual carries his own robust and yet unique way of moving (Individual Motor Signature – IMS; Hilt et al. 2016; Słowiński 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 a function of everyone’s individual movement style. Backed by this, we hypothesize that while observing others’ multi-joint actions, people build sensorimotor-based predictions by referencing what they see to the motor engrams of their own IMS.

To verify our hypothesis, we asked naive participants to first perform and then observe a whole-body reaching action which could be executed with numerous IMSs generally spread within a continuum between two “extreme” patterns (ankle and knee strategies; Hilt et al. 2016). After characterizing subjects’ own IMS during execution, we measured their sensorimotor recruitment (CSE) by administering single-pulse Transcranial Magnetic Stimulation (TMS) on their motor cortex while they observed an actor achieving the same goal by using the two “extreme” patterns of IMSs. CSE was measured from the cortical representation of the Tibialis Anterior muscle (TA) that shows a clearly dissociable pattern while executing the two IMSs. To exclude potential carry-over effects between action execution and observation, the same subjects were also tested several months later in the action observation task only.

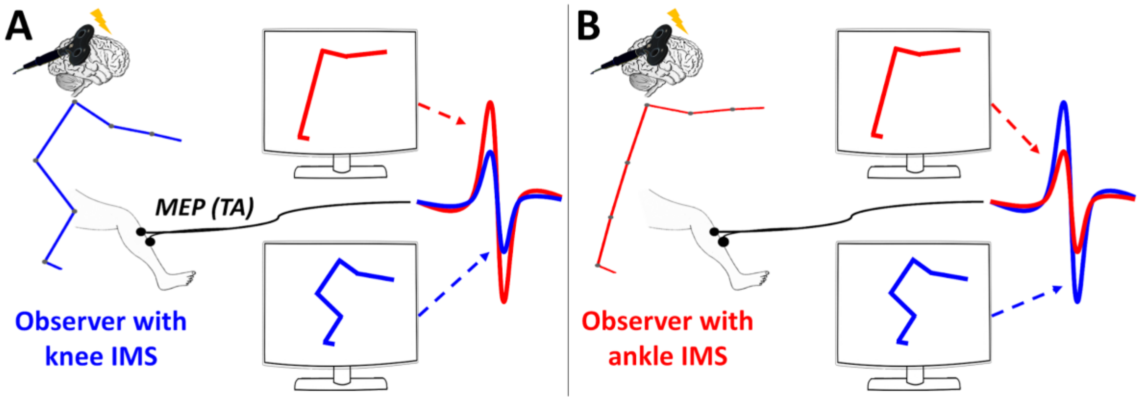


Figure 1: Illustration of the main results. MEPs amplitudes are depicted when observing knee (blue stick figure) or ankle (red stick figure) stimulus, for a subject that performed the knee (A) or the ankle (B) IMS in AE. Our results showed that corticospinal excitability was greater when actor and observer IMSs differ the most. These results agree with the predictive coding hypothesis that hypothesize the existence of a distance computation between observed movement and observer's IMS.

CSE was modulated at the single subject level according to the "distance" between actors' and observer's IMS: larger CSE modulations are associated with the observation of a more different IMS. This result is schematically illustrated in Figure 1 for two hypothetical subjects having extreme IMSs. Importantly, motor priming effects elicited by the action execution task can be excluded considering that the same pattern of results, in the same subjects, was shown several months later and in the absence of any action execution task.

Our results are at odds with a strictly simulative account of others' actions. Instead, the fact that sensorimotor activities during AO are shaped around a measure of distance between observed and own IMSs, agrees with the predictive coding framework. In this model, prior motor knowledge provides critical top-down signals that are integrated with bottom-up sensory-based processing (Friston 2010; Friston et al. 2011). To do so, a comparison between predicted (own IMS) and observed kinematic information (others' IMS) generates a prediction error signal that is used to update the representation of other's action.

Overall our data suggest that a greater uncertainty about other's action will call for a greater need of trustful predictions and consequently greater sensorimotor recruitment. In this context, the present study adds direct neurophysiological evidence that prediction errors are estimated by accessing IMS-related information. In fact, the many-to-many mapping problem in other's (multi-joint) action discrimination might be solved by accessing knowledge about IMSs. Indeed, the stability of IMSs (Słowiński et al. 2016; Coste et al. 2017) may reflect the implicit control and prioritization of a limited number of internal parameters during action planning and execution, partly solving the motor redundancy problem.

## References

Babiloni C, Babiloni F, Carducci F, et al (2002) Human Cortical Electroencephalography (EEG) Rhythms during the Observation of Simple Aimless Movements: A High-Resolution EEG Study. Neuroimage 17:559–572

Becchio C, Sartori L, Bulgheroni M, Castiello U. 2008. The case of Dr. Jekyll and Mr. Hyde: A kinematic study on social intention. Conscious Cogn. 17:557–564.

Bennequin D, Fuchs R, Berthoz A, Flash T. 2009. Movement Timing and Invariance Arise from Several Geometries. PLoS Comput Biol. 5:e1000426.

Bernstein NA (1967) The Coordination and Regulation of Movements, Pergamon P. Oxford

Berret B, Bonnetblanc F, Papaxanthis C, Pozzo T. 2009. Modular Control of Pointing beyond Arm ' s Length. J Neurosci. 29:191–205.

Blakemore SJ, Frith C. 2005. The role of motor contagion in the prediction of action. Neuropsychologia. 43:260–267.

Casile A, Dayan E, Caggiano V, Hendler T, Flash T, Giese MA. 2010. Neuronal encoding of human kinematic invariants during action observation. Cereb Cortex. 20:1647–1655.

Cochin S, Barthelemy C, Lejeune B, et al (1998) Perception of motion and qEEG activity in human adults. Electroencephalogr Clin Neurophysiol 107:287–295

Coste A, Slowinski P, Tsaneva-Atanasova K, Bardy BG, Marin L. 2017. Mapping Individual Postural Signatures. In: Weast-Knapp JA,, Pepping GJ, editors. Studies in Perception & Action XIV. Taylor & Francis.

D'Ausilio A, Bartoli E, Maffongelli L. 2015. Grasping synergies: A motor-control approach to the mirror neuron mechanism. Phys Life Rev. 12:91–103.

Dayan E, Casile A, Levit-Binnun N, Giese MA, Hendler T, Flash T. 2007. Neural representations of kinematic laws of motion: Evidence for action-perception coupling. Proc Natl Acad Sci. 104:20582–20587.

Decety J, Grèzes J, Costes N, et al (1997) Brain activity during observation of actions. Influence of action content and subject's strategy. Brain 120:1763–1777. doi: 10.1093/brain/120.10.1763

di Pellegrino G, Fadiga L, Fogassi L, et al (1992) Understanding motor events: a neurophysiological study. Exp Brain Res 91:176–180

Dinstein I (2008) Human Cortex : Reflections of Mirror Neurons. Curr Biol 18:956–959. doi: 10.1016/j.cub.2008.09.007

Fadiga L, Buccino G, Craighero L, et al (1998) Corticospinal excitability is specifically modulated by motor imagery: A magnetic stimulation study. Neuropsychologia 37:147–158. doi:10.1016/S0028-3932(98)00089-X

Fadiga L, Craighero L, Olivier E (2005) Human motor cortex excitability during the perception of others' action. Curr Opin Neurobiol 15:213–218. doi: 10.1016/j.conb.2005.03.013

Fadiga L, Fogassi L, Pavesi G, Rizzolatti G (1995) Motor facilitation during action observation: a magnetic stimulation study. J Neurophysiol 73:2608–2611

Falck-Ytter T, Gredebäck G, Von Hofsten C. 2006. Infants predict other people's action goals. Nat Neurosci. 9:878–879.

Flanagan JR, Johansson RS (2003) Action plans used in action observation. Lett to Nat 424:769–771. doi: 10.1038/nature01861

Fogassi L, Ferrari PF, Gesierich B, et al (2005) Parietal Lobe : From Action Organization to Intention Understanding. Science (80- ) 308:662–667. doi: 10.1126/science.1106138

Fox NA, Yoo KH, Bowman LC, et al (2016) Assessing human mirror activity With EEG mu rhythm: A meta-analysis. Psychol Bull 142:291–313. doi: 10.1037/bul0000031

Friston KJ. 2010. The free-energy principle: a unified brain theory? Nat Rev Neurosci. 11:127–138.

Friston KJ, Mattout J, Kilner JM. 2011. Action understanding and active inference. Biol Cybern. 104:137–160.

Gallese V, Fadiga L, Fogassi L, Rizzolatti G (2002) Action representation and the inferior parietal lobule. In: Common mechanisms in perception and action. pp 334–355

Hardwick RM, Caspers S, Eickhoff SB, Swinnen SP (2018) Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. Neurosci Biobehav Rev 94:31–44. doi: 10.1016/j.neubiorev.2018.08.003

Hilt PM, Bartoli E, Ferrari E, et al (2017) Action observation effects reflect the modular organization of the human motor system. Cortex 95:104–118. doi: 10.1016/j.cortex.2017.07.020

Hilt PM, Berret B, Papaxanthis C, et al (2016) Evidence for subjective values guiding posture and movement coordination in a free-endpoint whole-body reaching task. Sci Rep 6:23868. doi: 10.1038/srep23868

Horak FB, Nashner LM. 1986. Central programming of postural movements: adaptation to altered support-surface configurations. J Neurophysiol. 55:1369–1381.

Iacoboni M (2005) Neural mechanisms of imitation. Curr Opin Neurobiol 15:632–637. doi: 10.1016/j.conb.2005.10.010

Iacoboni M, Woods RP, Brass M, et al (1999) Cortical Mechanisms of Human Imitation. Sci New Ser 286:2526–2528. doi: 10.1038/020493a0

Kilner JM, Vargas C, Duval S, et al (2004) Motor activation prior to observation of a predicted movement. Nat Neurosci 7:1299–1301. doi: 10.1038/nn1355

Kilner JM. 2012. More than one pathway to action understanding. Trends Cogn Sci. 15:352–357.

Lapenta OM, Ferrari E, Boggio PS, et al (2018) Motor system recruitment during action observation: No correlation between mu-rhythm desynchronization and corticospinal excitability. PLoS One 13:1–15. doi: 10.1371/journal.pone.0207476

Mukamel R, Ekstrom AD, Kaplan J, et al (2010) Single-Neuron Responses in Humans during Execution and Observation of Actions. Curr Biol 20:750–756. doi: 10.1016/j.cub.2010.02.045

Naish KR, Houston-Price C, Bremner AJ, Holmes NP (2014) Effects of action observation on corticospinal excitability: Muscle specificity, direction, and timing of the mirror response. Neuropsychologia 64:331–348. doi: 10.1016/j.neuropsychologia.2014.09.034

Petrides M (2005) Lateral prefrontal cortex: Architectonic and functional organization. Philos Trans R Soc B Biol Sci 360:781–795. doi: 10.1098/rstb.2005.1631

Pineda JA (2008) Sensorimotor cortex as a critical component of an “extended” mirror neuron system: Does it solve the development, correspondence, and control problems in mirroring? Behav Brain Funct 4:1–16. doi: 10.1186/1744-9081-4-47

Rizzolatti G, Craighero L (2004) the Mirror-Neuron System. Annu Rev Neurosci 27:169–192. doi: 10.1146/annurev.neuro.27.070203.144230

Rizzolatti G, Fadiga L, Matelli M, et al (1996) Localization of grasp representations in humans by PET: 1. Observation versus execution. Exp Brain Res 111:246–252. doi: 10.1007/BF00227301

Rizzolatti G, Fogassi L, Gallese V (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. Nat Rev Neurosci 2:1–10. doi: 10.1038/35090060

Rizzolatti G, Sinigaglia C (2010) The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. Nat Rev Neurosci 11:264–274. doi: 10.1021/am4002502

Rizzolatti G, Sinigaglia C (2016) The mirror mechanism: a basic principle of brain function. Nat Rev Neurosci 17:757–765. doi: 10.1038/nrn.2016.135

Schütz-Bosbach S, Prinz W (2007) Prospective coding in event representation. Cogn Process 8:93–102. doi: 10.1007/s10339-007-0167-x

Sebanz N, Shiffrar M (2009) Detecting deception in a bluffing body: The role of expertise. Psychon Bull Rev 16:170–175. doi: 10.3758/PBR.16.1.170

Sebastiani V, de Pasquale F, Costantini M, et al (2014) Being an agent or an observer: Different spectral dynamics revealed by MEG. Neuroimage 102:717–728. doi: 10.1016/j.neuroimage.2014.08.031

Słowiński P, Zhai C, Alderisio F, Salesse R, Gueugnon M, Marin L, Bardy BG, di Bernardo M, Tsaneva-Atanasova K, 2016. Dynamic similarity promotes interpersonal coordination in joint action. J R Soc Interface. 13:20151093.

Soriano M, Cavallo A, D'Ausilio A, Becchio C, Fadiga L. 2018. Movement kinematics drive chain selection toward intention detection. Proc Natl Acad Sci U S A.

Turrella L, Pierno AC, Tubaldi F, Castiello U (2009) Mirror neurons in humans : Consisting or confounding evidence ? Brain Lang 108:10–21. doi: 10.1016/j.bandl.2007.11.002

## 2.1.6 Investigate singularity in ellipses drawing

The goal of the experiment is to identify a high level feature able to characterize movement of different people. More in detail, this feature will be detected in writing movements over drawings of a geometrical shape such as an ellipse. Since the movement of each person are different from another one, this high level feature is called singularity  and can be represent the first point to assess motion signature of people.

However, this is only the first step of this experiment and in the future, also the perception of the movement will be investigated.

Measuring singularity can contribute to many application fields, from the clinical to entertainment and customer applications.

In the first experiment we have focused on motor signature aspect: people try several times the drawing of the same ellipse under several conditions. More in detail, these conditions characterize the hand that draws the ellipse (both right and left hand are investigated = 2) and the drawing speed (slow, normal and fast = 3) for a total of 2x3 conditions. Moreover, for each condition the participant repeat the experiment 10 times drawing at each repletion of the experiment 7 ellipses. Therefore, the cardinality of the datasets is obtained by 14 (the number of participants) x 2 (hands) x 3 (speeds) x 10 (trials) x 7 (ellipses) = 5880 available ellipses.

Data are collected using a graphic table and the available raw data for each ellipse are the positions on the screen (x and y position), the velocity, the curvature and the pressure on the tablet.

From the hierarchy presented in the dataset, we focused our analysis identifying two different scenarios able to determine the behavior of the model trained. These 2 scenarios, providing different sensibilities on data used in training set, can be used to understand and estimate the algorithm behavior on unseen or new data. The 2 scenarios are:

- Leave-One-Hand-Out (LOHO): the learning set of our classifier is made up of all people of the dataset except the information coming from one hand of the tested person.

This scenario is interesting because one hand is the dominant one and can be very relevant in the final classification.

- Leave-One-Speed-Out (LOSO): the learning set of our classifier is made up of all people, all hands and all speeds in the dataset, except one speed coming from the tested person. In this case, the learning set contains more information respect to the previous scenario and this will be reflected in higher recognition results.

To study the singularity, machine learning models are used to classify the drawings of people. In particular, both shallow models and deep learning ones are exploited. As results, we will show the goodness of a multiple temporal scales approach respect traditional ones that not taking into account this intrinsic aspect of the human movement.

## Traditional machine learning model

Traditional machine learning models can provide stable and robust results but in order to improve their recognition performances, features that can describe and catch the behavior over time are needed.

In this experiment we have performed this operation, which is called Feature Extraction or Feature Engineering (FE), segmenting ellipses in several ways in order to find which representation can improve the final outcome. In particular, we have used the following splits:

a) More straight parts and more linear parts are considered separately (Figure 1.(a));

b) First half end second half of the ellipse are considered separately (Figure 1.(b));

c) Each curve and each linear parts are considered separately (Figure 1.(c));

d) Respect the case (c), we further considered each linear part divided in according to the diagonal (Figure 1.(d);

e) All the previous splits are considered at the same time.

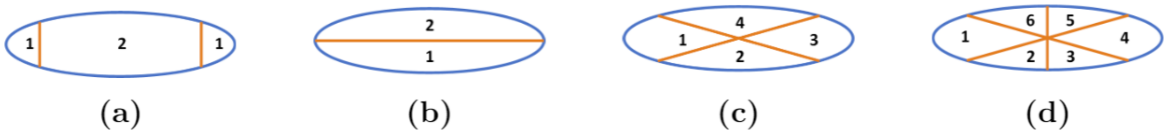


Figure 1: The ellipse criteria of segmentation.

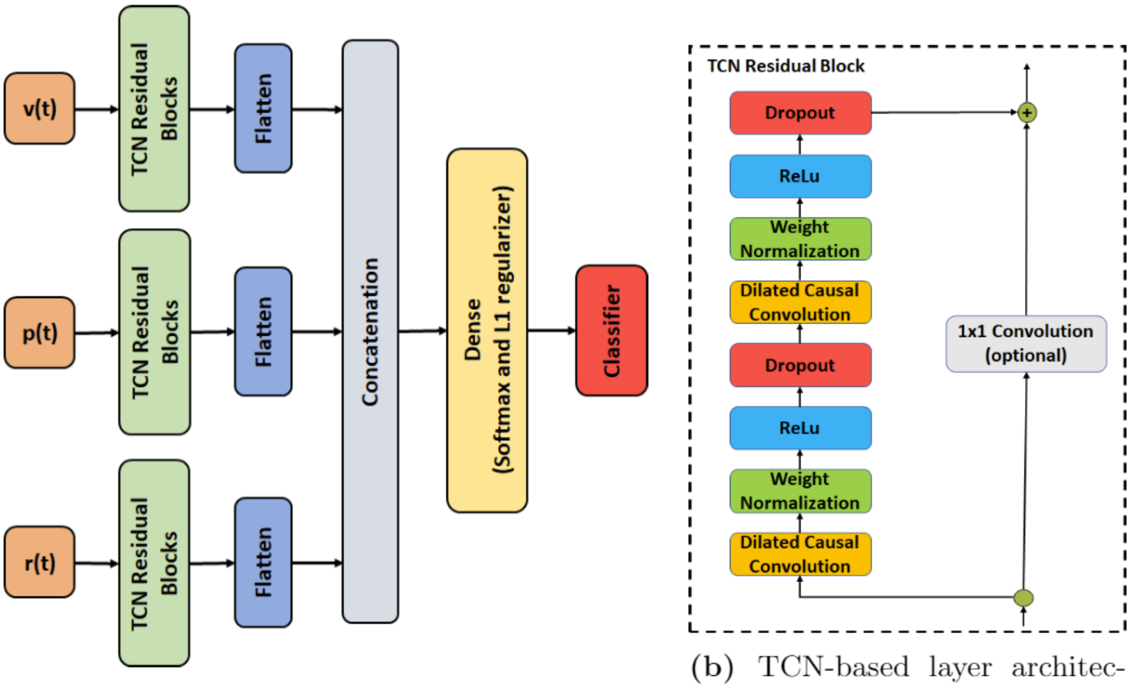
As a consequence, for each criteria of segmentation, we extracted statistical feature on the different sections of the ellipses.
A powerful algorithm, both in terms of theoretical properties and practical effectiveness (Fernández-Delgado et al 2014, Wainberg et al. 2016), for classification is Random Forest (RF) developed in (Breiman et al., 2001) for the first time. RF is composed of the union of multiple Decision Trees (Rokach et al. 2008). Compared to DTs, RF introduces an additional degree of randomness due to the introduction of a bootstrap phase.

Finally, a Feature Ranking (FR) step is computed in order to discover the most relevant section of an ellipse. Once a model is built, it is often required to understand how this model exploits, combine, and extract information in order to understand if the learning process has also cognitive meaning, namely it is able to capture the underline phenomena and does not just capture spurious correlation (Calude et al., 2017; Guyon et al., 2003)  by comparing the knowledge of the experts with the information learned by the models. FR therefore represents a fundamental phase of model checking and verification, since it should generate results consistent with the available knowledge of the phenomena under exam provided by the experts.
FR methods based on RF are one of the most effective FR techniques as shown in many research (Genuer et al., 2010; Saeys et al., 2008). Several measures and approaches are available for FR in RF. One method is based on the Permutation Test combined with the Mean Decrease in Accuracy (MDA) metric, where the importance of each feature is estimated by removing the association between the feature and outcome of the model. For this purpose, the values of the features are randomly permuted (Good et al.; 2013) and the resulting increase in error is measured. In this way also the influence of the correlated features is also removed. Note that, in our case, as a feature we do not intend a particular engineered feature but a particular ellipse section (e.g.the first section when split= 6, the second curve sections when split=4, etc.).

## Deep Learning models

A parallel approach is related to the use of Deep Learning (DL) models. As we know, these architectures are automatically able to extract the best set of features from raw data. This implies that the feature engineering step and therefore, the sections split, are not needed anymore.

To better understand multiple time-scales we can follow different approaches in DL. Architectures such as Clockwork-RNN (Koutnik et al., 2014) or Multi-LSTM (Liu et al., 2015) are designed to automatically detect multi-time scales information. Other models such as LSTM (Hochreiter et al., 1997), simply RNN, Multi-Layer Perceptron (MLP), are not directly thought to handle this problem but can provide excellent results if properly used. To overcome this limitation we decided to rely on TCN residual blocks (Bai et al., 2018, Lee et al., 2017) which is capable to learn different temporal scales for each raw input time series. The proposed architecture is reported in Figure. The peculiarities of the proposed Deep Multi Scale Models architecture based on TCN, which is visible in Figure 2, are mainly three: first the convolutions in the architecture are causal, namely there is not information leakage from future to past, second the architecture can handle different sequence lengths and map it to an output sequence of the same length as the LSTM, and finally is able to handle long effective history.



(a) High level representation of the architecture.

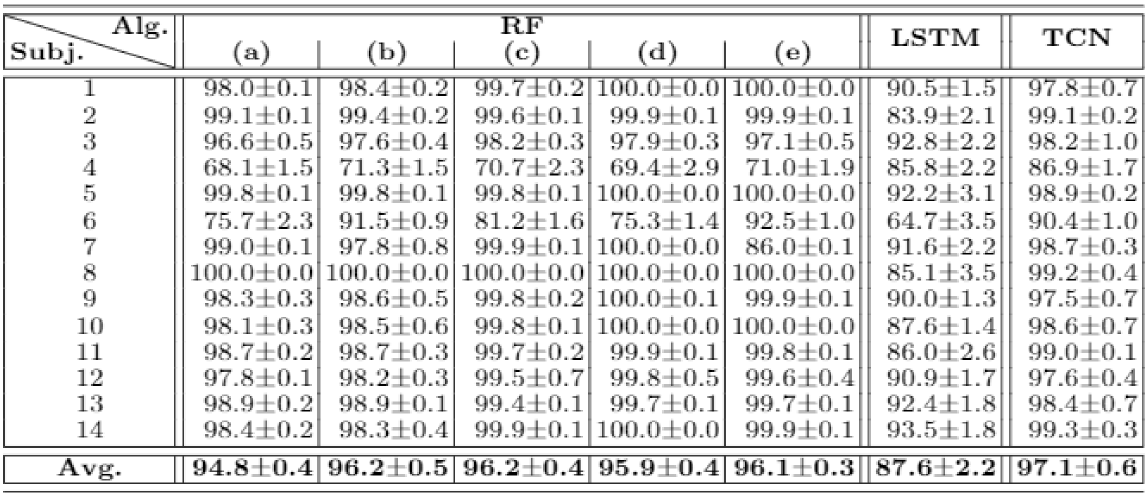
(b) TCN-based layer architecture details.

Figure 2: The proposed Deep Multi Scale Models architecture based on TCN.

To provide a complete comparison respect to the model that belong to the state-of-the-art in Machine Learning, we analyzed both LSTM and Multiple Temporal-Scales TCN. Moreover, as for the shallow models, to understand what parts of the time series (velocity, radius or pressure) mostly contribute to the decision, we decides to exploit the Gradient weighted Class Activation Mapping (Grad-CAM) () techniques on top of the learned model. Grad-CAM allow to easily visualize the most important part of the input time series since it extends traditional class activation maps and can be applied to a broader variety of architectures. In fact, the result of Grad-CAM is a localization map that highlights key sections in an input for a given class, providing insights on where neural networks focus their attention.

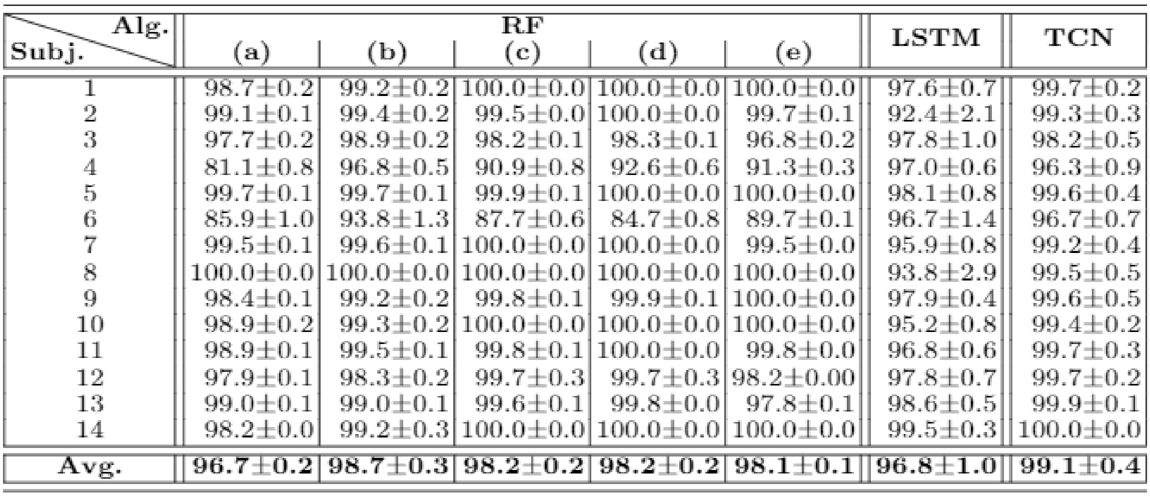
## Recognition performances in LOHO and LOSO

Let discuss now the recognition performances obtained in the two scenarios we identified. Table 1. reports the percentage of accuracy, in the LOHO and LOSO scenarios respectively, when exploiting RF (with the different criteria of segmentation of the ellipses described in Section Traditional Machine Learning), LSTM and TCN for each of the 14 subjects together with the average across the subjects.



| Alg. / Subj. | RF (a) | (b) | (c) | (d) | (e) | LSTM | TCN |
|---|---|---|---|---|---|---|---|
| 1 | 98.0±0.1 | 98.4±0.2 | 99.7±0.2 | 100.0±0.0 | 100.0±0.0 | 90.5±1.5 | 97.8±0.7 |
| 2 | 99.1±0.1 | 99.4±0.2 | 99.6±0.1 | 99.9±0.1 | 99.9±0.1 | 83.9±2.1 | 99.1±0.2 |
| 3 | 96.6±0.5 | 97.6±0.4 | 98.2±0.3 | 97.9±0.3 | 97.1±0.5 | 92.8±2.2 | 98.2±1.0 |
| 4 | 68.1±1.5 | 71.3±1.5 | 70.7±2.3 | 69.4±2.9 | 71.0±1.9 | 85.8±2.2 | 86.9±1.7 |
| 5 | 99.8±0.1 | 99.8±0.1 | 99.8±0.1 | 100.0±0.0 | 100.0±0.0 | 92.2±3.1 | 98.9±0.2 |
| 6 | 75.7±2.3 | 91.5±0.9 | 81.2±1.6 | 75.3±1.4 | 92.5±1.0 | 64.7±3.5 | 90.4±1.0 |
| 7 | 99.0±0.1 | 97.8±0.8 | 99.9±0.1 | 100.0±0.0 | 86.0±0.1 | 91.6±2.2 | 98.7±0.3 |
| 8 | 100.0±0.0 | 100.0±0.0 | 100.0±0.0 | 100.0±0.0 | 100.0±0.0 | 85.1±3.5 | 99.2±0.4 |
| 9 | 98.3±0.3 | 98.6±0.5 | 99.8±0.2 | 100.0±0.1 | 99.9±0.1 | 90.0±1.3 | 97.5±0.7 |
| 10 | 98.1±0.3 | 98.5±0.6 | 99.8±0.1 | 100.0±0.0 | 100.0±0.0 | 87.6±1.4 | 98.6±0.7 |
| 11 | 98.7±0.2 | 98.7±0.3 | 99.7±0.2 | 99.9±0.1 | 99.8±0.1 | 86.0±2.6 | 99.0±0.1 |
| 12 | 97.8±0.1 | 98.2±0.3 | 99.5±0.7 | 99.8±0.5 | 99.6±0.4 | 90.9±1.7 | 97.6±0.4 |
| 13 | 98.9±0.2 | 98.9±0.1 | 99.4±0.1 | 99.7±0.1 | 99.7±0.1 | 92.4±1.8 | 98.4±0.7 |
| 14 | 98.4±0.2 | 98.3±0.4 | 99.9±0.1 | 100.0±0.0 | 99.9±0.1 | 93.5±1.8 | 99.3±0.3 |
| **Avg.** | **94.8±0.4** | **96.2±0.5** | **96.2±0.4** | **95.9±0.4** | **96.1±0.3** | **87.6±2.2** | **97.1±0.6** |

(a) LOHO



| Alg. / Subj. | RF (a) | (b) | (c) | (d) | (e) | LSTM | TCN |
|---|---|---|---|---|---|---|---|
| 1 | 98.7±0.2 | 99.2±0.2 | 100.0±0.0 | 100.0±0.0 | 100.0±0.0 | 97.6±0.7 | 99.7±0.2 |
| 2 | 99.1±0.1 | 99.4±0.2 | 99.5±0.0 | 100.0±0.0 | 99.7±0.1 | 92.4±2.1 | 99.3±0.3 |
| 3 | 97.7±0.2 | 98.9±0.2 | 98.2±0.1 | 98.3±0.1 | 96.8±0.2 | 97.8±1.0 | 98.2±0.5 |
| 4 | 81.1±0.8 | 96.8±0.5 | 90.9±0.8 | 92.6±0.6 | 91.3±0.3 | 97.0±0.6 | 96.3±0.9 |
| 5 | 99.7±0.1 | 99.7±0.1 | 99.9±0.1 | 100.0±0.0 | 100.0±0.0 | 98.1±0.8 | 99.6±0.4 |
| 6 | 85.9±1.0 | 93.8±1.3 | 87.7±0.6 | 84.7±0.8 | 89.7±0.1 | 96.7±1.4 | 96.7±0.7 |
| 7 | 99.5±0.1 | 99.6±0.1 | 100.0±0.0 | 100.0±0.0 | 99.5±0.0 | 95.9±0.8 | 99.2±0.4 |
| 8 | 100.0±0.0 | 100.0±0.0 | 100.0±0.0 | 100.0±0.0 | 100.0±0.0 | 93.8±2.9 | 99.5±0.5 |
| 9 | 98.4±0.1 | 99.2±0.2 | 99.8±0.1 | 99.9±0.1 | 100.0±0.0 | 97.9±0.4 | 99.6±0.5 |
| 10 | 98.9±0.2 | 99.3±0.2 | 100.0±0.0 | 100.0±0.0 | 100.0±0.0 | 95.2±0.8 | 99.4±0.2 |
| 11 | 98.9±0.1 | 99.5±0.1 | 99.8±0.1 | 100.0±0.0 | 99.8±0.0 | 96.8±0.6 | 99.7±0.3 |
| 12 | 97.9±0.1 | 98.3±0.2 | 99.7±0.3 | 99.7±0.3 | 98.2±0.00 | 97.8±0.7 | 99.7±0.2 |
| 13 | 99.0±0.1 | 99.0±0.1 | 99.6±0.1 | 99.8±0.0 | 97.8±0.1 | 98.6±0.5 | 99.9±0.1 |
| 14 | 98.2±0.0 | 99.2±0.3 | 100.0±0.0 | 100.0±0.0 | 100.0±0.0 | 99.5±0.3 | 100.0±0.0 |
| **Avg.** | **96.7±0.2** | **98.7±0.3** | **98.2±0.2** | **98.2±0.2** | **98.1±0.1** | **96.8±1.0** | **99.1±0.4** |

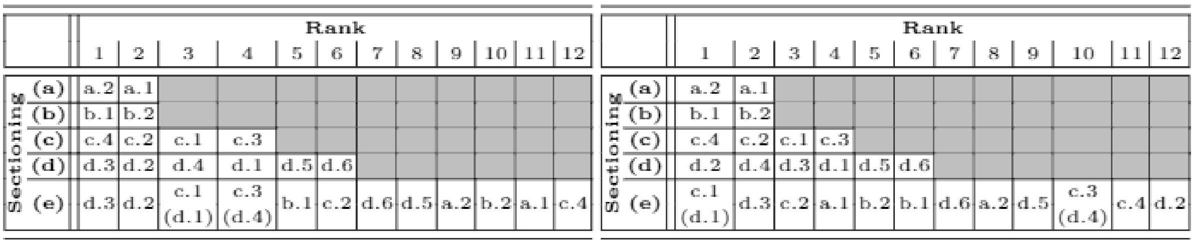
(b) LOSO

Table 1: Percentage of accuracy when exploiting RF (with the different criteria of segmentation of the ellipses, LSTM and TCN for each of the 14 subjects together with the average across the subjects.

Table 1. allows to observe that:

- as one might expect performances on LOSO are generally higher than the ones on LOHO for all subjects and algorithms. This is the natural consequence of the fact that in LOHO we are asking a more complex extrapolation capability to the algorithms;
- TCN consistently outperform RF and LSTM in all scenarios while demonstrating also consistent performance between subjects;
- RF is quite competitive and outperform, for some subjects, also TCN. Nevertheless, for some subjects, performance are quite poor;
- RF in case (a) and (b) performs quite well. These results indicate that segmenting too little or too much the ellipse is not a good solution while putting all the possible segmentations, as in case (e), does not guarantees optimal performance. In fact, these segmentation designed to capture multiple time scales are, by construction, fixed and not customized for the specific problem. The TCN-based architecture, instead, actually learns the correct time scale to focus on;
- LSTM, as expected, is the algorithm which demonstrates the lowest performance. This is due to the fact that its ability to capture different time scales is too limited.

In order to better understand what the different RF and TCN models actually learned from the data, Table 2. reports the sections ranking. The letter indicates the sectioning and the number indicates the specific section so note that c.1 is the same as d.1 and c.3 is the same as d.4. performed with RF in the different sectioning scenarios and Figure 3. reports the attention maps of TCN, averaged across subjects, for pressure p(t), velocity v(t) and radius r(t) for both LOHO and LOSO scenarios.



|  |  | Rank |  |  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Sectioning | (a) | a.2 | a.1 |  |  |  |  |  |  |  |  |  |  |
|  | (b) | b.1 | b.2 |  |  |  |  |  |  |  |  |  |  |
|  | (c) | c.4 | c.2 | c.1 | c.3 |  |  |  |  |  |  |  |  |
|  | (d) | d.3 | d.2 | d.4 | d.1 | d.5 | d.6 |  |  |  |  |  |  |
|  | (e) | d.3 | d.2 | c.1<br>(d.1) | c.3<br>(d.4) | b.1 | c.2 | d.6 | d.5 | a.2 | b.2 | a.1 | c.4 |

(a) LOHO

|  |  | Rank |  |  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Sectioning | (a) | a.2 | a.1 |  |  |  |  |  |  |  |  |  |  |
|  | (b) | b.1 | b.2 |  |  |  |  |  |  |  |  |  |  |
|  | (c) | c.4 | c.2 | c.1 | c.3 |  |  |  |  |  |  |  |  |
|  | (d) | d.2 | d.4 | d.3 | d.1 | d.5 | d.6 |  |  |  |  |  |  |
|  | (e) | c.1<br>(d.1) | d.3 | c.2 | a.1 | b.2 | b.1 | d.6 | a.2 | d.5 | c.3<br>(d.4) | c.4 | d.2 |

(b) LOSO

Table 2: Sections ranking performed with RF in the different sectioning scenarios for both LOHO and LOSO scenarios.

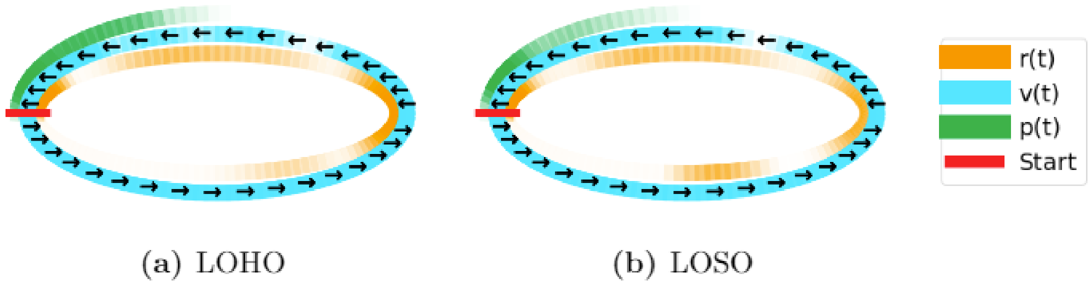


Figure 3: Attention maps of TCN, averaged across subjects, for $v(t)$, $r(t)$, and $p(t)$ for both LOHO and LOSO scenarios. The more intense is the colour, the more important is the particular part of the input time series.

Table 2. and Figure 3. allow to observe that:

- as one might expect, the sections in the two scenarios are not exactly the same since they try to extrapolate with respect to different information (hand and speed). When using shallow models (i.e., RF) for sectioning (a), (b), and (c) sections maintain the same importance in both LOHO and LOSO scenarios while for sectioning (d) and (e) the ranking is quite different. When using deep models (i.e., TCN), instead, only for v(t) the attention map remain similar for both LOHO and LOSO scenarios;
- for both LOHO and LOSO scenarios, shallow models identify as the most informative sections those who are closer to the initial part of the drawing in all the analysed sectioning criteria. On the other hand, deep models generally find the final parts of the drawing as most informative. This shows how different is the perception of the two models. The shallow ones focus on the “preparation” of the movement, while the deep ones focus more on the “completion” of the movement. The deep model, in this case, perceives the movement in a way which seems more similar to a human: human beings tend to become more confident in labelling a movement when it tends to be completed;
- shallow models primarily focus on more “linear” sections with respect to the more “curved ones”. The opposite happens for deep models. Also in this case, deep model perception is more similar to human one: human tends to distinguish movements based on the most complex parts;

- finally note that shallow models tend to focus on sections based on the particular choice of the sectioning criteria and are not able to perceive and define their one way of understanding the movement. Deep models, instead, by construction are able to do so defining the attention maps based on the particular problem and defining implicitly their own sectioning criteria then being able to perceive the different time scales of the movement.

## References

Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271(2018)

Breiman, L.: Random forests. Machine Learning45(1), 5–32 (2001)

Calude, C.S., Longo, G.: The deluge of spurious correlations in big data. Foundations of science22(3), 595–612 (2017)

Fernandez-Delgado, M., Cernadas, E., Barro, S., Amorim, D.: Do we need hundreds of classifiers to solve real world classification problems? The Journal of Machine LearningResearch15(1), 3133–3181 (2014)

Genuer, R., Poggi, J.M., Tuleau-Malot, C.: Variable selection using random forests. Pattern Recognition Letters31(14), 2225–2236 (2010)

Good, P.: Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer Science & Business Media (2013)

Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research3(Mar), 1157–1182 (2003)

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Koutnik, J., Greff, K., Gomez, F., & Schmidhuber, J. (2014). A clockwork rnn. arXiv preprint arXiv:1402.3511.

Lee, S.M., Yoon, S.M., Cho, H.: Human activity recognition from accelerometerdata using convolutional neural network. In: IEEE international conference on bigdata and smart computing (2017)

Liu, P., Qiu, X., Chen, X., Wu, S., & Huang, X. J. (2015, September). Multi-timescale long short-term memory neural network for modelling sentences and documents. In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 2326-2335).

Masci, J., Meier, U., Cireşan, D., & Schmidhuber, J. (2011, June). Stacked convolutional auto-encoders for hierarchical feature extraction. In International conference on artificial neural networks (pp. 52-59). Springer, Berlin, Heidelberg.

Rokach, L., Maimon, O.Z.: Data Mining with Decision Trees: Theory and Applications, vol. 69. World Scientific (2008)

Saeys, Y., Abeel, T., Van de Peer, Y.: Robust feature selection using ensemble feature selection techniques. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases (2008)

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In:IEEE international conference on computer vision (2017)

Wainberg, M., Alipanahi, B., Frey, B.J.: Are random forests truly the best classifiers? The Journal of Machine Learning Research17(1), 3837–3841 (2016)

## 2.1.8 Perception of the origin of full body human movement and its propagation

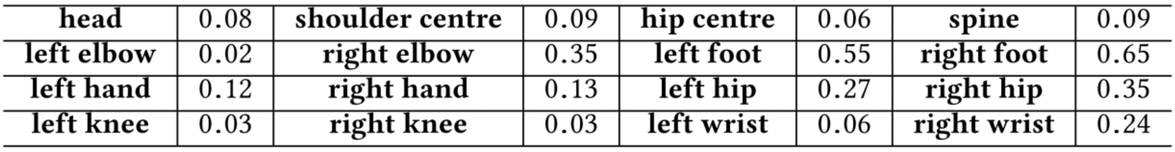
We further developed and extended the approach proposed in (Kolykhalova et al., 2020, Gnecco et al., 2020). In particular, we investigated a more complex skeletal structure, in which each cluster of joints is associated to a specific joint in the simpler 20-joint skeletal structure. This allows one to analyze movement at a finer interacting spatio-temporal scale in a multiple-scale approach. Another development regards using a vector of movement-related features to compute the Shapley value, in order to get a comparison with the results obtained in Kolykhalova e al. (2020), where only speed was used as a feature.

Incorporate multiple temporal scales. For example, one could look at a fast temporal scale at a first step of the analysis of the origin of movement, then at a slower temporal scale where one can analyse the origin of movement at a higher level.

Applying the methodology to analyse the emergence of the origin of movement when two persons or small groups are involved in the movement itself is the next step. We are currently running collaboration with EuroMov looking into multi-person scenarios.

A recent paper (Matthiopoulou et al., 2020) including preliminary results on some of these extension of the theory was presented in a joint UNIGE-EuroMov paper at the ACM ICMI 2020 EnTimeMent Workshop. The additional features investigated therein (beside speed) were tangential acceleration and angular momentum. Moreover, we investigated therein the loss in information associated with the reduction of the original 62-joint skeletal structure to a 20-joint one, after a manual clustering of joints performed on the original structure. An additional feature, called “mass distribution”, was also defined. Loosely speaking, such feature quantifies how much each cluster of joints behaves as a rigid body. A small coefficient of variation of that feature is then associated with a small loss of information when moving from the more complex skeletal structure to the less complex one. As an example, Table 1 illustrates, for a specific fragment, the coefficients of variation obtained for some clusters of joints (excluding the ones made by a single joint).

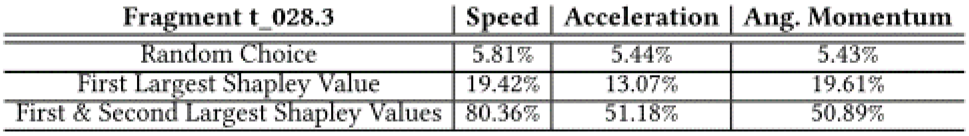
Table 1



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **head** | 0.08 | **shoulder centre** | 0.09 | **hip centre** | 0.06 | **spine** | 0.09 |
| **left elbow** | 0.02 | **right elbow** | 0.35 | **left foot** | 0.55 | **right foot** | 0.65 |
| **left hand** | 0.12 | **right hand** | 0.13 | **left hip** | 0.27 | **right hip** | 0.35 |
| **left knee** | 0.03 | **right knee** | 0.03 | **left wrist** | 0.06 | **right wrist** | 0.24 |

Some comparison of the proposed approach when it was applied to different movement related features (speed, tangential acceleration, angular momentum) are reported in Table 2. A comparison with random choice is also reported. The table shows that, for the specific fragment analyzed, the proposed method worked much better than random choice (whose expected average classification accuracy was 5%, being the reduced skeletal structure made by 20 joints). Moreover, the largest agreement between the labeled origin of movement and the one predicted by the proposed method was achieved when speed was the adopted feature, and both the first and second largest Shapley values were produced as output by the method (i.e., the origin of movement was evaluated to be identified correctly when at least one of the two joints coincided with it).

Table 2



| **Fragment t_028.3** | **Speed** | **Acceleration** | **Ang. Momentum** |
|---|---|---|---|
| Random Choice | 5.81% | 5.44% | 5.43% |
| First Largest Shapley Value | 19.42% | 13.07% | 19.61% |
| First & Second Largest Shapley Values | 80.36% | 51.18% | 50.89% |

A theoretical analysis of the effects of large-dimensional classification was developed in (Kůrková and Sanguineti, 2021) from a probabilistic point of view.

As regards the acquisition of new data regarding the dyadic scenario, a set of recordings was created, based on a pair of participants exchanging a ball launched using both hands at a distance of about 3m from each other. Two different balls were used, one light (about 100g), and one heavier (about 2Kg). We defined three conditions for the repeated exchange of the ball to each other: fair (the launcher tries to facilitate the receiver), hit/hurt (the launcher tries to hit the receiver), misleading (the launcher tries to reduce the success of the receiver by misleading actions).

In the dyadic scenario several complementary research questions emerge, for instance : (i) investigating the origin of movement in the sender and in the receiver, (ii) determining at which time scale the origin of movement is perceived, (iii) evaluating the synchronization and anticipatory behaviour between sender and receiver, (iv) evaluationg leader-follower behaviour in the different emotions, (v) and how the origin of movement contributes to capture those different emotions. We are currently working in these research directions.

## References

Gnecco, G., Sanguineti, M., Camurri, A., Bardy, B., and Mottet, D. (2021). Comparing Features for Detecting the Origin of Movement based on a Graph-Theoretical Cooperative Game Model. To be presented at the International Conference Optimization and Decision Science (ODS 2021)

Kůrková, V., and Sanguineti, M. (2021). Correlations of Random Classifiers on Large Data Sets. Soft Computing, to appear. DOI: 10.1007/s00500-021-05938-4

Matthiopoulou, O., Bardy, B., Gnecco, G., Mottet, D., Sanguineti, M., and Camurri, A. (2020). A Computational Method to Automatically Detect the Perceived Origin of Full-Body Human Movement and its Propagation. In Companion Publication of the 2020 ACM International Conference on Multimodal Interaction (ICMI 2020), pp.449-453.

Matthiopoulou, O., Gnecco, G., Sanguineti, M., Camurri, A., Bardy, B., and Mottet, D. (2020). Detecting the Perceived Origin of Full-Body Human Movement via Shapley Values Games on Graphs. International Conference Optimization and Decision Science (ODS 2020)

Kolykhalova, K., Gnecco, G., Sanguineti, M., Volpe, G., and Camurri, A. (2020) Automated analysis of the origin of movement: An approach based on cooperative games on graphs, IEEE Transactions on Human-Machine Systems, vol. 50, pp. 550 – 560. DOI: 10.1109/THMS.2020.3016085

## 2.1.11 Computation-Based Feature Representation of Body Expressions in the Human Brain

**For a full description please see**: Poyo Solanas, Marta, Maarten Vaessen, and Beatrice de Gelder. "Computation-Based Feature Representation of Body Expressions in the Human Brain." *Cerebral Cortex* (2020).

Humans and other primate species are experts at recognizing body expressions. To understand the underlying perceptual mechanisms, we computed postural and kinematic features from affective whole-body movement videos and related them to brain processes. Using representational similarity and multivoxel pattern analyses, we showed systematic relations between computation-based body features and brain activity. Our results revealed that postural rather than kinematic features reflect the affective category of the body movements. The feature limb contraction showed a central contribution in fearful body expression perception, differentially represented in action observation, motor preparaton, and affect coding regions, including the amygdala. The posterior superior temporal sulcus differentiated fearful from other affective categories using limb contraction rather than kinematics. The extrastriate body area and fusiform body area also showed greater tuning to postural features. The discovery of midlevel body feature encoding in the brain moves affective neuroscience beyond research on high-level emotion representations and provides insights in the perceptual features that possibly drive automatic emotion perception.

It is widely agreed that humans and other primate species are experts at recognizing emotion and intention from face and body expressions (de Gelder 2006; Giese and Rizzolatti 2015). The central importance of nonverbal communication across many social species suggests that the brain is equipped for rapid and accurate face and body movement perception; yet, the mechanisms underlying this ability are still largely unclear. Previous research on face and body expressions has predominantly searched for brain correlates of symbolic emotion categories (Lindquist et al. 2012; Kirby and Robinson 2017), disregarding the visual features that drive movement and emotion perception (e.g., kinematic and postural body features). This is in part due to the fact that methods for fine-grained description of body movements were not yet available. This study used computational descriptions of body expressions to investigate which features drive emotion and body perception and how they are encoded in the brain.

Previous behavioral and computational studies have provided some indications about relevant features of body posture and movement, and their relation to emotional expressions (De Meijer 1989; Wallbott 1998; Roether et al. 2009; Kleinsmith and Bianchi-Berthouze 2012; Piana et al. 2014; Patwardhan 2017). Some important postural features have been identified, including elbow flexion, associated with the expression of anger, and head inclination, typically observed for sadness (Wallbott 1998; Coulson 2004; Vaessen et al. 2018). Other form-related features that have been investigated are the vertical extension of the body (e.g., upper limbs remain low for sadness but high for happiness), the directionality of the movement (e.g., angry bodies are usually accompanied by a forward movement), symmetry (e.g., the movement of the upper limbs tends to be symmetrical when experiencing joy), and the amount of lateral opening of the body (e.g., hands are close to the body during fear and sadness while extended in happiness) (Kleinsmith and Bianchi-Berthouze 2012).

A central, yet unanswered, question is the relation between candidate features and brain processes. There is sparse evidence in the literature on how particular features may be related to brain processes. One classical proposal is the two-stream model of visual processing with two separate brain pathways for form and movement information (Vaina et al. 1990; Giese and Poggio 2003; Milner and Goodale 2006, 2008). From the primary visual cortex, the dorsal stream leads to the parietal lobe and is specialized in localizing objects in space, processing motion signals and in the visual-spatial guidance of actions. The ventral stream leads to the temporal lobe and is responsible for visual form processing and object recognition. Two areas in this pathway have been identified that sustain a certain level of specialization in the processing of whole bodies and body parts: the extrastriate body area (EBA) in the medial occipital cortex, and the fusiform body area (FBA) in the fusiform gyrus (Downing et al. 2001; Peelen and Downing 2005; Schwarzlose et al. 2005). However, their respective functions are not yet clear and it is also not clear how they, alone or together, contribute to body expression perception. In addition, body shape and movement elicit a widespread neural response beyond the visual analysis of body features in body-category selective areas (de Gelder 2006; Van den Stock et al. 2011), triggering processes related to their affective content, the conveyed action and for the preparation of an appropriate behavioral response (de Gelder et al. 2004; Van den Stock et al. 2011). The present study is the first effort to discover which specific postural and kinematic features could be computed from affective whole-body movement videos and be related to brain responses. By means of representational similarity multivoxel pattern analysis techniques, we investigated whether the (dis)similarity of body posture and kinematics between different emotional categories could explain neural responses to body expressions in and beyond body-selective regions.

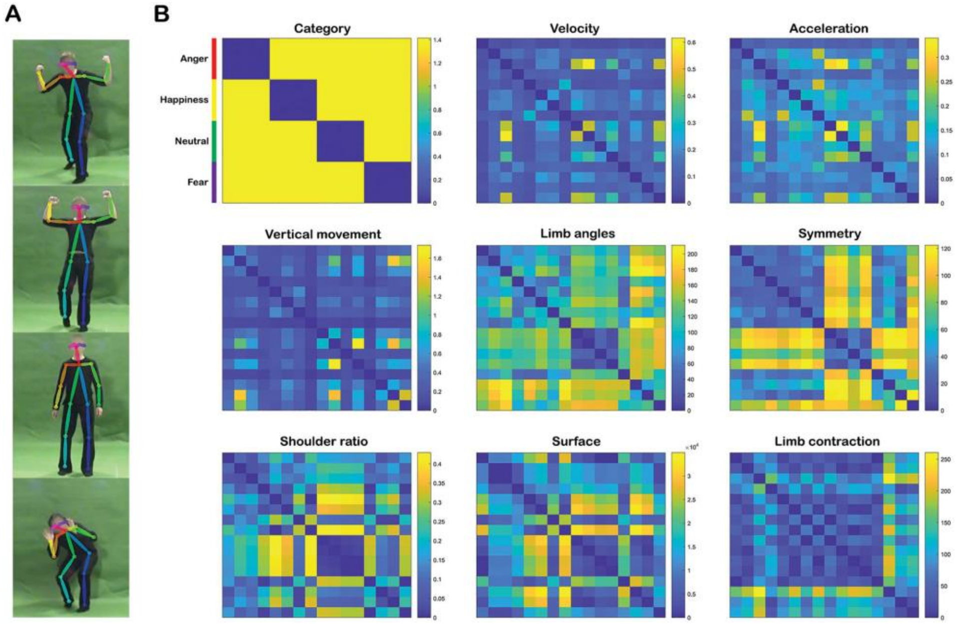


Figure 7: Representational dissimilarity matrices of the kinematic and postural features. (A) Examples of frames from the different affective movement videos with the OpenPose skeleton. Note that participants were shown the videos without the OpenPose skeleton; (B) The RDMs represent pairwise comparisons between the 16 stimuli with regard to the kinematic (i.e., velocity, acceleration, and vertical movement) and postural features (i.e., limb angles, symmetry, shoulder ratio, surface, and limb contraction) averaged over time. The dissimilarity measure reflects Euclidean distance, with blue indicating high similarity and yellow

high dissimilarity. Color lines in the upper left corner indicate the organization of the RDMs with respect to the emotional category (anger: red; happiness: yellow; neutral: green; fear: purple) of the video stimuli.

We aimed at investigating whether the (dis)similarity of body posture and kinematics between different emotional categories could explain the neural response of brain regions involved in body processing. For this purpose, several areas were defined as ROI and their neural RDMs were computed and correlated to the emotional and feature RDMs. The ROIs included occipito-temporal areas that have previously shown a certain level of body specificity (three ROIs: FBA, EBA, and pSTS) (Downing et al. 2001; Peelen and Downing 2005; Schwarzlose et al. 2005; Kontaris et al. 2009; Vangeneugden et al. 2014), parietal and temporal areas thought to be implicated in attention and action observation (six ROIs: V7/3a, SPOC, SMG, pIPS, mIPS, and aIPS) (Culham and Valyear 2006; Grafton and Hamilton 2007; Corbetta et al. 2008; Caspers et al. 2010), and frontal areas involved in action observation and other higher cognitive functions (six ROIs: PMv, PMd, SMA, pre-SMA, inferior frontal,and frontal regions) (Grafton and Hamilton 2007; Caspers et al. 2010). See Figure 8 for the full results.

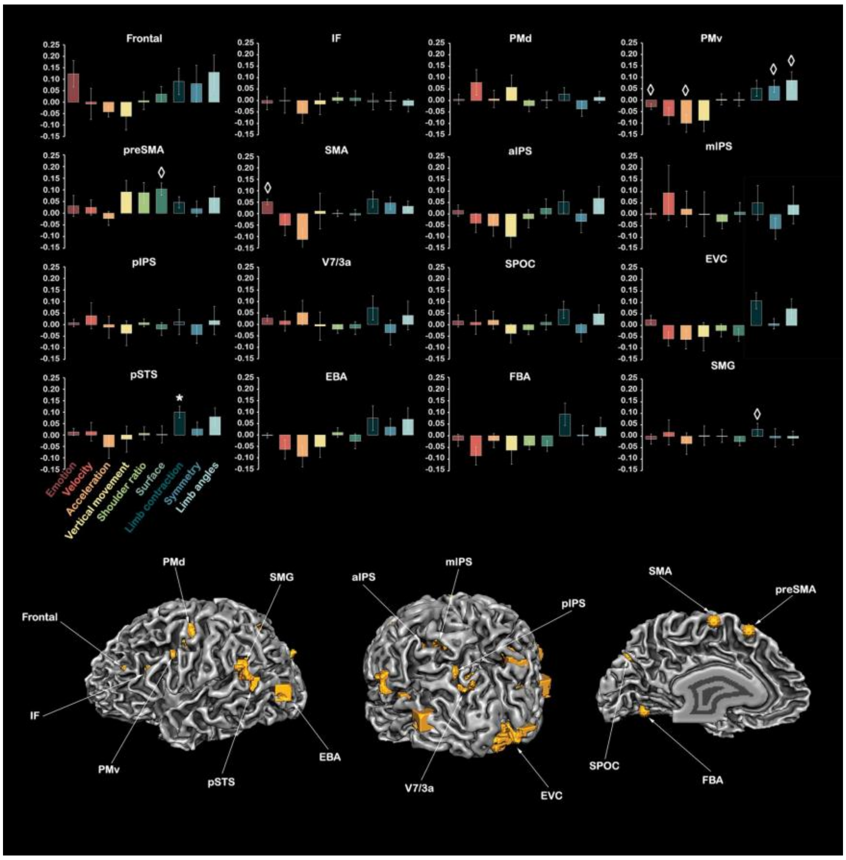


Figure 8: Average Spearman's rank correlation across participants between the kinematic/postural feature RDMs and each ROI matrix. Kinematic features include velocity, acceleration, and vertical movement. Postural features comprise shoulder ratio, surface, limb contraction, symmetry, and limb angles. Positive r values indicate that a high (dis)similarity between a stimulus pair in the feature RDM also has a high (dis)similarity in the neural representation. A negative correlation means that a low (dis)similarity between two stimuli at the feature level would have a higher (dis)similarity in the neural representation. Asterisks and rhombi indicate significant correlations after BHFDR correction and correlations that presented significant uncorrected P-values, respectively (one sample t-test against 0, two-tailed). The error bars denote the standard error of the mean (SEM). Order or relationships across ROIs are not assumed here. Abbreviations: EBA, extrastriate body area; EVC, early visual cortex; FBA, fusiform body area; IF, inferior frontal cortex; IPS, intraparietal sulcus; p, posterior; m, middle; a, anterior; PMd, dorsal premotor cortex; PMv, ventral premotor cortex; pre-SMA, presupplementary motor area; pSTS, posterior superior temporal sulcus; SMA, supplementary motor area; SMG, supramarginal gyrus; SPOC, superior parietal occipital cortex.

We also investigated whether (dis)similarities in body posture and kinematics between different emotional categories could explain the neural response at the whole-brain level. The computed feature RDMs were compared with the multivoxel dissimilarity fMRI patterns by means of searchlight RSA. The velocity RDM was positively correlated to inferior frontal sulcus and precentral gyrus. Negative main effects for acceleration were found in middle temporal, superior frontal, and postcentral sulci while no positive main effects were observed for this feature. Vertical movement correlated positively with cingulate gyrus, whereas negatively to the frontomarginal and middle temporal gyri. With respect to postural features, limb angles showed a positive main effect in anterior insula and pSTS. Several areas negatively correlated to symmetry in the inferior and middle occipital gyri, precuneus, isthmus, anterior calcarine, intraparietal, and cingulate sulcus. Shoulder ratio negatively correlated to anterior insula, frontal operculum, putamen, ACC, middle frontal gyrus, cingular insular sulcus, claustrum, internal capsule, and parahippocampal gyrus. Surface showed main negative effects in posterior orbital gyrus, thalamus, anterior perforated substance, ACC, inferior and superior frontal sulci, putamen, and internal capsule. Only positive correlations to limb contraction were found in intraparietal sulcus, anterior insula, caudate nucleus, amygdala, superior frontal sulcus and gyrus, precuneus, posterior orbital gyrus, ACC, superior temporal gyrus, inferior precentral sulcus, and SMG (see Figure 9).

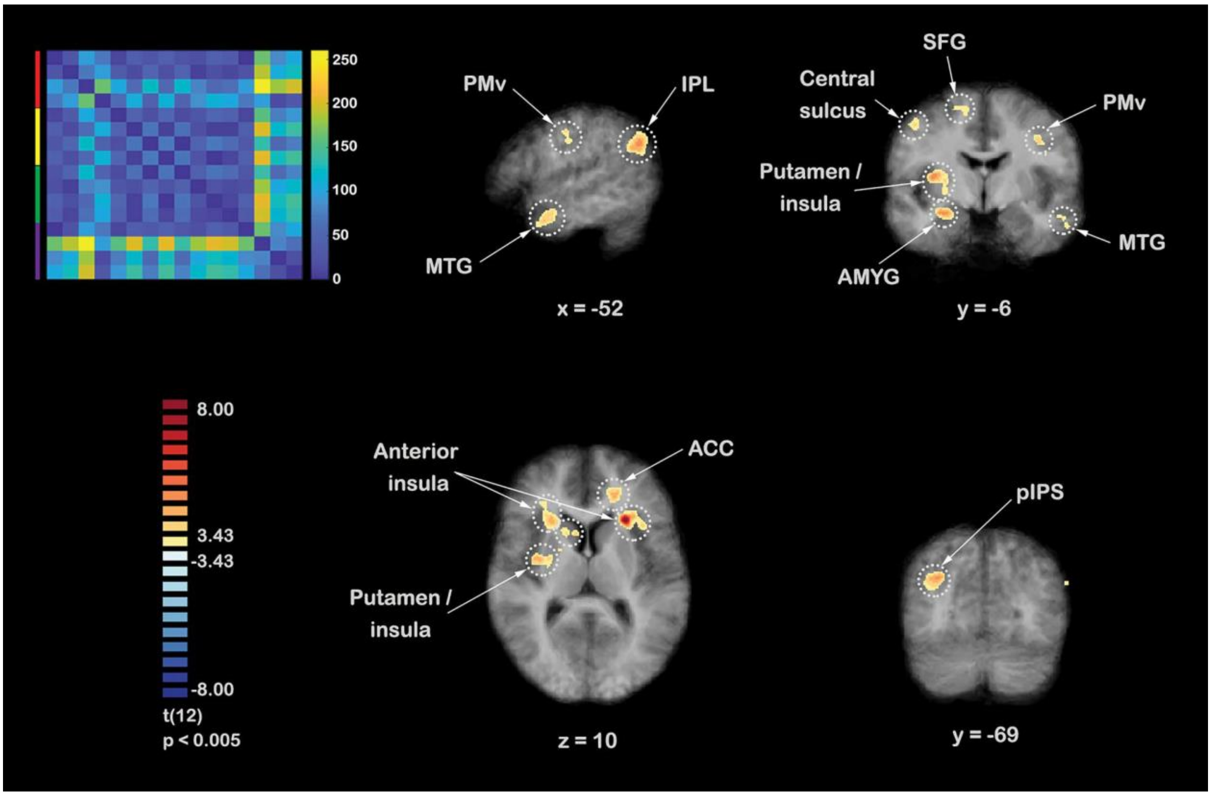


Figure 9: Clusters resulting from the searchlight RSA of the postural feature of limb contraction. The multivoxel fMRI dissimilarity matrices were correlated to the limb contraction RDM (upper left corner). The limb contraction RDM represents pairwise comparisons between the 16 stimuli with regard to limb contraction information averaged over time. The dissimilarity measure reflects Euclidean distance, with blue indicating high similarity and yellow high dissimilarity. Color lines indicate the organization of the RDM with respect to the emotional category (anger: red; happiness: yellow; neutral: green; fear: purple) of the video stimuli. Spearman's rank correlation was used to correlate the limb contraction RDM to the multivoxel fMRI dissimilarity matrices. The resulting maps were z-transformed for each participant. Subsequently, a group-level one-sample t-test against 0was performed (two-tailed, cluster size corrected with Monte-Carlo simulation, alpha level = 0.05, initial P = 0.005, numbers of iterations = 5000). See Supplementary Table R5 in Supplementary Results for more details on location and statistical values of the clusters. Abbreviations: ACC, anterior cingulate cortex; AMYG, amygdala; IPL, inferior parietal lobule; MTG,middle temporal gyrus; pIPS, posterior intraparietal sulcus; PMv, ventral premotor cortex; SFG, superior frontal gyrus.

The present study investigated the mechanisms underlying body expression perception by measuring the brain representation of critical features of body movement and posture. Our results reveal six major findings. First, computationally defined features are systematically related to distributed brain areas. Second, postural rather than kinematic features reflect the affective category structure of the body movements. Limb angles and symmetry were important for differentiating neutral from emotional body movements. Limb angles and especially limb contraction were particularly relevant for distinguishing fear from other body expressions. These two features were represented in several regions including affective, action observation and motor preparation networks. Third, the pSTS differentiated fearful from other affective categories using limb contraction rather than kinematics, despite this area being known for its involvement in biological motion processing. Fourth, EBA and FBA also showed greater tuning to postural features. Although the pattern of feature representation in these areas was similar, the stimuli representation in EBA was very dissimilar to that of FBA, possibly reflecting their different roles in body processing. Fifth, kinematic and postural feature processing was not segregated into dorsal and ventral streams, with the exception of one feature: velocity. Finally, the brain representation of emotional categories showed a distributed pattern.

By investigating mid level feature processes, this study moves the field of affective neuroscience forward, providing insights into the perceptual features that possibly drive automatic emotion perception. Features at this visual computational level may only partly overlap with feature descriptions used in everyday descriptions of body expressions (Poyo Solanas et al. 2020). Nevertheless, it is important to be aware of the limitations of our findings. For instance, the features defined here were selected due to their relevance in the literature because no feature-based and biologically plausible computational model of naturalistic body expressions is available (Giese and Poggio 2003; Serre 2014). We expect that future studies will also use larger and more diverse stimulus sets with a wider range of affective states and a larger participant sample, also looking into dyadic interactions.

## References

Caspers S, Zilles K, Laird AR, Eickhoff SB. 2010.ALE meta-analysis of action observation and imitation in the human brain. Neuroimage. 50:1148–1167.

Coulson M. 2004. Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. J Nonverbal Behav. 28:117–139.

de Gelder B, Snyder J, Greve D, Gerard G, Hadjikhani N. 2004. Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. PNAS. 101:16701–16706.

de Gelder B. 2006. Towards the neurobiology of emotional body language. Nat Rev Neurosci. 7:242.

De Meijer M. 1989. The contribution of general features of body movement to the attribution of emotions. J Nonverbal Behav. 13:247–268.

Downing PE, Jiang Y, Shuman M, Kanwisher N. 2001. A cortical area selective for visual processing of the human body. Science. 293:2470–2473.

Giese MA, Poggio T. 2003. Neural mechanisms for the recognition of biological movements. Nat Rev Neurosci. 4:179–192.

Giese MA, Rizzolatti G. 2015. Neural and computational mechanisms of action processing: interaction between visual and motor representations. Neuron. 88:167–180.

Grafton ST, Hamilton AF. 2007. Evidence for a distributed hierarchy of action representation in the brain. Hum Mov Sci. 26:590–616.

Kirby LA, Robinson JL. 2017. Affective mapping: an activation likelihood estimation (ALE) meta-analysis. Brain Cogn. 118:137–148.

Kleinsmith A, Bianchi-Berthouze N. 2012. Affective body expression perception and recognition: a survey. IEEE T Affect Comput. 4:15–33.

Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF. 2012. The brain basis of emotion: a meta-analytic review. Behav Brain Sci. 35:121

Milner D, Goodale MA. 2006. The visual brain in action.Oxford (UK): Oxford University Press.

Patwardhan A. 2017. Three-dimensional, kinematic, human Behavioral pattern-based features for multimodal emotion recognition. Multimodal Technol Interact. 1:19.

Peelen MV, Downing PE. 2005. Selectivity for the human body in the fusiform gyrus. J Neurophysiol. 93:603–608.

Piana S, Stagliano A, Odone F,Verri A,Camurri A. 2014. Real-time automatic emotion recognition from body gestures. arXiv preprint arXiv:1402.5047

Poyo Solanas M, Vaessen M, de Gelder B. 2020. The role of computational and subjective features in emotional body expressions. Sci Rep. 10:1–13.

Roether CL, Omlor L, Christensen A, Giese MA. 2009. Critical features for the perception of emotion from gait.J Vis. 9:15–15.

Schwarzlose RF, Baker CI, Kanwisher N. 2005. Separate face and body selectivity on the fusiform gyrus. J Neurosci. 25:11055–11059.

Serre T. 2014. Hierarchical models of the visual system. In: Jung R, Jaeger D, editors. Encyclopedia of computational neuroscience. New York: Springer.

Vaessen M, Abassi E, Mancini M, Camurri A, de Gelder B. 2018. Computational feature analysis of body movements reveals hierarchical brain organization. Cereb Cortex. 1:10.

Vaina LM, Lemay M, Bienfang DC, Choi AY, Nakayama K. 1990. Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: a case study. Vis Neurosci. 5:353–369.

Van den Stock J, Tamietto M, Sorger B, Pichon S, Grézes J, de Gelder B. 2011. Cortico-subcortical visual, somatosensory, and motor activations for perceiving dynamic whole-body emotional expressions with and without striate cortex (V1). PNAS. 108:16188–16193.

Wallbott HG. 1998. Bodily expression of emotion. Eur J Soc Psychol. 28:879–896.

## 2.1.14 Automatic Detection in the Context of Movement with Chronic Pain based on Novel Multiple-Timescales Machine Learning Architectures

In this section, we present the results of the exploration of our novel machine learning architectures (*Body Attention Net*, i.e. BANet; *Movement in Multiple Time*, i.e. MiMT; *Global Workspace Network*, i.e. GWN; and *Hierarchical Human Activity Recognition and Protective Behaviour Detection,* i.e. HAR-PBD*, model*) on the problem of automatic detection of pain and related behaviour from body movement. Please D3.1 for descriptions of the BANet and MiMT which we also refer to as Multi Time neural network (MTNN) or MultiLevNN. D1.6 and 1.7 provide descriptions of the GWN and Hierarchical HAR-PBD model respectively.

Pain behaviour assessment is an important movement analysis problem in the context of chronic pain physical activities (Cook et al. 2013; Keefe and Block 1982). Automating the assessment of pain behaviour could enable less burdensome, objective measurement as well as open up the opportunity to provide real-time tailoring (e.g. via movement sonification) which fosters engagement of a person with chronic pain with valued physical activities. For example, a sonification framework based on pain behaviour assessment could aim to call the attention of a person with pain to their unhelpful strategies for performing feared or painful movements (Olugbade et al. 2019). Previous bodily-expressed pain behaviour detection studies have focused on classification at single timescales. For example, (Aung et al. 2016) modelled the duration (as a proportion) of pain behaviour in a movement instance based on a simple fusion of motion capture and muscle activity data. Pain level itself could be valuable to assess automatically for the purpose of enabling helpful pacing of everyday physical activities based on the understanding of how pain drives underactivity and overactivity (Olugbade et al. 2019).
For all 4 investigations (on the BANet, MiMT, GWN, and Hierarchical HAR-PBD respectively), we used the EmoPain dataset (Aung et al. 2016) which contains 3D positions for 26 full-body joints, 13 full-body angles derived from these, and muscle

activity data for 4 upper and lower back locations of people with chronic pain and healthy control participants. The data were captured while the participants performed 8 exercise movements (sit-to-stand, stand-to-sit, bend, reach forward, walk, sitting, standing, one leg raised while standing). People with chronic pain provided self-reports of pain after each exercise type on a scale of 0 to 10. Four clinicians further provided annotations for the exercise instances of this group of participants, for 6 pain behaviours (guarding/stiffness, hesitation, bracing/support, abrupt action, limping, rubbing, stimulation) in continuous time and as discrete values 0 for 'not present' and 1 as 'present'.

### Study 1: Weighted Fusion of Time and Anatomical Region with The BANet

A contribution of the BANet is its importance weighting of time for each movement dimension (e.g. hip angle) and further weighting of each dimension overall. Further details can be found in the peer-reviewed publication of the study.

Focusing on 5 (sit-to-stand, stand-to-sit, reach forward, bend, one leg raised while standing) of the 8 exercise types in EmoPain dataset, this study was based on the 13 full-body joint angles of the dataset and the angular energies computed from them. Each joint angle is derived from the 3D positions of three consecutive joints while the corresponding joint energy is the square of the change in the angle with respect to the previous timestep. The angle and energy sequences for each exercise type participant were segmented based on fixed window with length = 3 seconds and overlapping ratio = 0.75 based on findings in (Wang et al. 2019), within each exercise instance. Zeroes were used to pad segments at the end of exercise instance and less than the window length. Two data augmentation techniques were then applied to duplicates of these segments to increase the data size, i.e. the number of segments. The first of these adds normalized Gaussian noise (standard deviation = 0.05, 0.1, and 0.15) to the duplicate (based on Wang et al. 2019). In the second, randomly selected (probability = 0.05, 0.1, and 0.15) angles and energies in the duplicate are dropped, i.e. set to 0 (based on Um et al. 2017). The augmentation resulted in 18,653 segments. The ground truth for each segment was set to 'protective behaviour present' if at least 2 of the 4 raters rated any of the pain behaviours as present for half of the segment length and 'absent' otherwise. There were 11,373 *protective behaviour absent* segments and 7,280 *protective behaviour present* segments.

We evaluated the performance of the BANet on automatic discrimination between protective behaviour absent and present classes using leave-one-subject-out cross-validation. To understand the value of approach used in the BANet, we compared its performance on automatic detection of protective behaviour based on these data with the performance of 4 variants of the BANet (BANet-compatibility, BANet-dense, BANet-time-only, BANet-body-only). We also compared the BANet with 3 architectures (bidirectional Long Short-Term Memory neural network i.e. LSTMNN, convolutional LSTMNN, stacked LSTMNN) which are similar to it but do not include weighting, i.e. the machine learning attention mechanism which gives the BANet its name. Table 4 gives an overview of all 8 architectures explored in this study and the hyperparameters used in training the respective models. The Adam optimiser and learning rate of 0.003 was used in all cases.

Table 4: The BANet and 7 peer machine learning architectures that we compared it to.

| Architecture | Attention (i.e. weighting) across time | Attention (i.e. weighting) across joint (angle) | Layers Types [number of layers, number of units] | Training batch size |
|---|---|---|---|---|
| BANet | Yes | Yes but after time attention | 1. LSTM [3, 8]<br>2. 1x1 convolution and softmax (time attention)<br>3. fully connected [2, 8] and softmax (joints attention)<br>4. fully connected [1, 2] and softmax | 40 |

| | | | | |
|---|---|---|---|---|
| BANet-compatible | Yes but after joints attention | Yes | 1. LSTM [3, 8]<br>2. fully connected [2, 8] and softmax (joints attention)<br>3. 1x1 convolution and softmax (time attention)<br>4. fully connected [1, 2] and softmax | 40 |
| BANet-dense | Yes | Yes but after time attention | 1. LSTM [3, 8]<br>2. fully connected [1, 8] and softmax (time attention)<br>3. fully connected [2, 8] and softmax (joints attention)<br>4. fully connected [1, 2] and softmax | 40 |
| BANet-time-only | Yes | No | 1. LSTM [3, 8]<br>2. 1x1 convolution and softmax (time attention)<br>3. fully connected [1, 2] and softmax | 40 |
| BANet-body-only | No | Yes | 1. LSTM [3, 8]<br>2. fully connected [2, 8] and softmax (joints attention)<br>3. fully connected [1, 2] and softmax | 40 |
| Bidirectional LSTMNN | No | No | 1. bidirectional LSTM [1, 14] and dropout probability of 0.5 | 40 |
| Stacked LSTMNN | No | No | 1. LSTM [3, 28] and dropout probability of 0.5 | 20 |
| Convolutional LSTMNN | No | No | 1. 1x10 convolution, 28 LSTM units, and max pooling | 50 |

The performance of the BANet can be seen in Table 5 in comparison with the other 7 architectures. As can be seen in the table, the BANet outperforms all of the 7 architectures. A paired t test over cross-validation folds, with Bonferroni correction, showed that this is statistically significant ($p<0.05$) for every comparison architecture except the BANet-time-only and BANet-body-only F(3.072, 89.099)=15.612, $\mu^2$= 0.519); the significance for the bidirectional LSTM was marginal.

Table 5: Mean F1 score and accuracy of the BANet and comparison architectures (in bold is the best performance and * is used to indicate comparison architectures which performed significantly worse, $p<0.05$, than the BANet).

| Architecture | Mean F1 Score | Accuracy | Number of trainable parameters |
|---|---|---|---|
| BANet | 0.8440 | 0.8688 | 2,131 |
| BANet-compatible | 0.5720* | 0.6630 | 6,204 |
| BANet-dense | 0.7890* | 0.8167 | 65,430 |
| BANet-time-only | 0.7580 | 0.8060 | 1,767 |
| BANet-body-only | 0.8310 | 0.8670 | 2,023 |
| Bidirectional LSTMNN | 0.8040 | 0.8460 | 14,282 |
| Stacked LSTMNN | 0.8120* | 0.8534 | 18,986 |
| Convolutional LSTMNN | 0.7370* | 0.8059 | 40,940 |

Another advantage of the attention weighting of the BANet is that it allows analysis of both temporal and anatomical segment relevance. Figure 9 shows boxplots of the distribution of attention scores (i.e. importance weights) for each joint angle (and its energy) per exercise type. It can be seen that there is a wider distribution of attention scores for the participants with chronic pain, particularly in the exercise segments with protective behaviour absent, compared with the healthy participants. This suggests strong salience of a few antomical segments above the others, perhaps in terms of distinction in timescale, with protective behaviour. The sample plots of temporal attention scores per joint angle and energy in Figure 10 showing larger differences in the timelines for the different joint angles supports this theory.

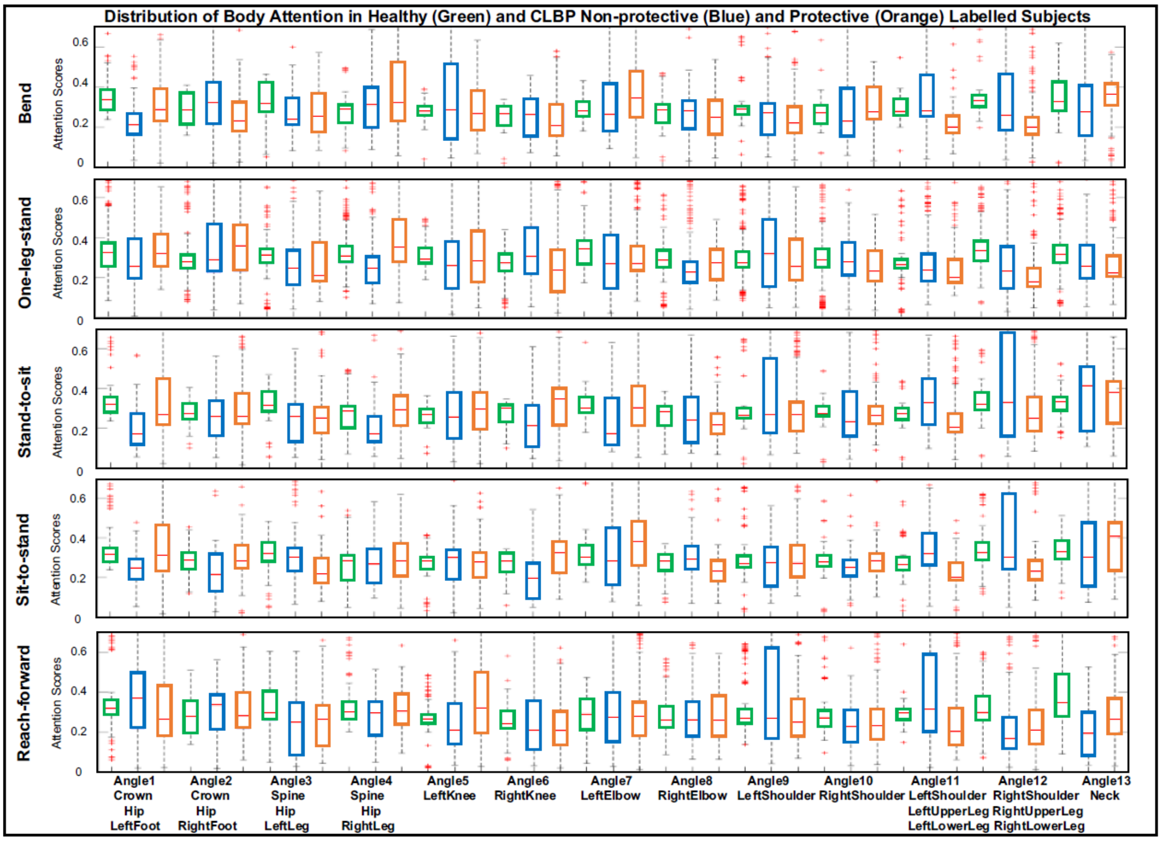


Figure 9: Distribution of attention scores for each joint angle (and its energy) per exercise type. The plots show healthy participants in green, participants with chronic pain and protective behaviour absent in blue, and participants with chronic pain and protective behaviour present in orange.

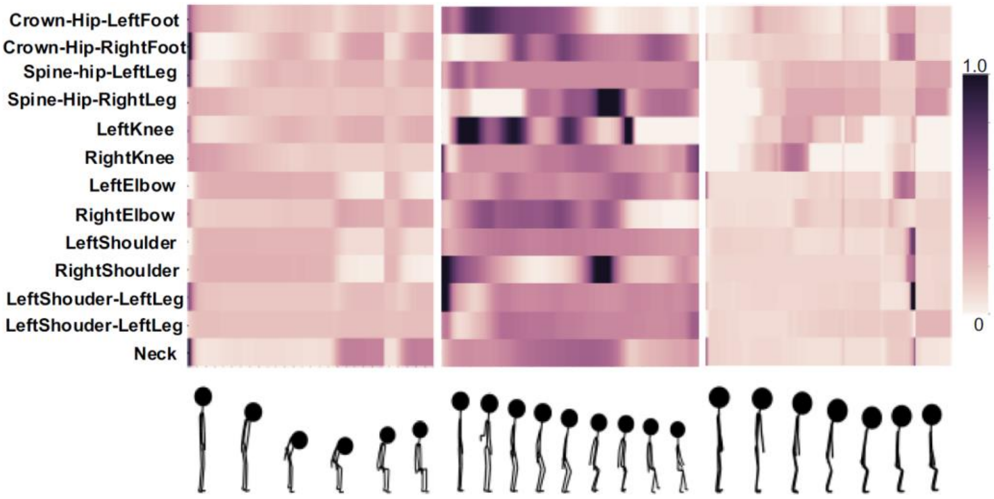


Figure 10: Sample plots of temporal attention per joint angle (and its energy) in a stand-to-sit exercise instance for a healthy participant (left) and two participants with chronic pain (middle and right).

We propose a novel neural network architecture named BANet which performs weighted fusion of movement time and anatomical regions. This approach outperforms similar architectures without explicit weights in fusion, with weights only  for time or anatomical region but not both, or with the weighted fusion of anatomical regions before time.

Analysis of these weights, which are learnt by the network based on data, suggests stronger differences in timescales of anatomical segments during anomalous movement behaviour. First, this highlights that multiple timescales occur not just over time itself but also across the different degrees of freedom of movement. We have developed a movement sonification framework that aims to apply multi-dimensionality of time (attention time and the different times of each degrees of freedom of movement) in chronic pain scenarios. On one hand, this could be used to provide self-awareness (attention) in real time to a person with chronic pain about how they are moving. On the other hand, it could serve as to augment an observer's (the person with pain themselves or a clinician) visual assessment of movement. More details about the sonification framework is reported in D3.1.  Second, it raises questions about how much the network attention weighting tells us about the timescales involved in the interpretation of movement behaviour by the clinicians who provided protective behaviour labels. We are carrying out further analyses of the attention scores to understand what they imply in this respect. Further, we are additionally conducting an observation study aimed at finding implicit models of pain and movement that physiotherapists use to make clinical observations and interventions.

### Study 2: Using The MiMT to Learn Multiple Timescales of Pain Behaviour Labels based on Movement Dimensions with Multiple Timescales

The MiMT models body movement at multiple timescales particularly accounting for independence-cum-coordination between multiple anatomical segments similar to the BANet but also accounting for different timescales of movement interpretation (at level of a single time step and at the level of multiple timesteps). Further details can be found in the peer-reviewed publication of the study.

While the joint angles used for Study 1 have the advantage of being location invariant, we chose to use the 3D full-body positions of the EmoPain dataset in Study 2 because they characterise movement execution in a more intuitive way. We excluded eight of the 26 joints (left and right fingertips, ankles, heels, and toes) in our use of the positional data in this study due to the higher level of noise in their position estimates. To minimise the dimensionality of the data, we additionally excluded the crown joint given that the remaining joints include the head and neck. This resulted in 17 full-body joints. We segmented exercise instances in the EmoPain dataset (except the *walking* exercises and for participants with chronic pain alone) using overlapping 3-second windows based on (Wang et al. 2019) (overlap = 0.25 seconds). The label for a frame (timestep) in a segment was set as *of guarding* if at least two raters labelled guarding behaviour as present at that frame, otherwise the label was set as *not of guarding*. The label for a segment (multiple timesteps) was set as *guarding behaviour* if all the frames in the segment are *of guarding* label, *not guarding behaviour* if all the frames are *not of guarding*, and *mixed* otherwise. We used data augmentation to increase the minority classes at the segment level (*guarding behaviour*, *mixed*) by creating mirror duplicates across permutations of the three axes (based on Olugbade et al. 2020) as well as translated, scaled up/down duplicates. This resulted in 17,185 and 1,394 instances respectively for the training and validation sets.

We evaluated the MiMT on automatic discrimination between of guarding and not of guarding at the frame level and between guarding, not guarding, and mixed classes at the segment level. This evaluation was based on hold-out validation where the subject sets in the training, validation, and test sets are mutually exclusive. To understand the value of the approach of the MiMT (separate but shared time encoding and multiple timescales of the same label), we compared its performance with 3 architectures derived by ablation of the MiMT (MiMT-single-input-time, MiMT-frame-output-time-only, MiMT-segment-output-time-only). Table 6 provides an overview of the differences between the architectures. The time encoder of the MiMT (and the comparison architectures) was based on 3 LSTM layers each with 3 units. Single LSTM and fully connected layers each with 15 units were used for the classifier with additional global average pooling and sigmoid activation for the frame level output and a single layer LSTM and softmax activation after further multiplication with the time encoder output for the segment level output.  Each model was trained with the Adam optimizer at learning rate and batch size of 0.005 and 200 respectively.

Table 6: An overview of the architectures compared with the MiMT.

| Architecture | Separate but shared time encoding of the input | Frame level output | Segment level output |
|---|---|---|---|
| MiMT | Yes | Yes | Yes |
| MiMT-single-input- time | No | Yes | Yes |
| MiMT-frame-output-time-only | Yes | Yes | No |
| MiMT-segment-output-time-only | Yes | No | Yes |

Table 7 shows the performance of the MiMT. As can be seen in the table, the MiMT performs much better than chance level detection (0.5 for the frame label, 0.33 for the segment label). The MiMT further outperforms its three variants suggesting that combination of both the separate but shared time encoding for the input and the multiple output timescales is efficacious.

Table 7: Mean F1 score of the MiMT and comparison architectures (in bold is the best performance).

| | Mean F1 score | |
|---|---|---|
| Architecture | Frame label | Window label |
| MiMT | 0.63 | 0.46 |
| MiMT-single-input- time | 0.50 | 0.34 |
| MiMT-frame-output-time-only | 0.59 | - |
| MiMT-window-output-time-only | - | 0.33 |

Figure 11 shows two example plots of the activations for each separate (but shared) time encoding. To maximise contrast, we only sampled every 20th frame in these plots. Each band for each group of segments represents the activation for one of the three units of the encoder.  Comparing the bands for the lower left and right limb groups of segments clearly show coordination between the two groups of segment yet there are differences in changes in the activations over time further highlighting that different degrees of freedom have different timescales that have moments of synchronization.

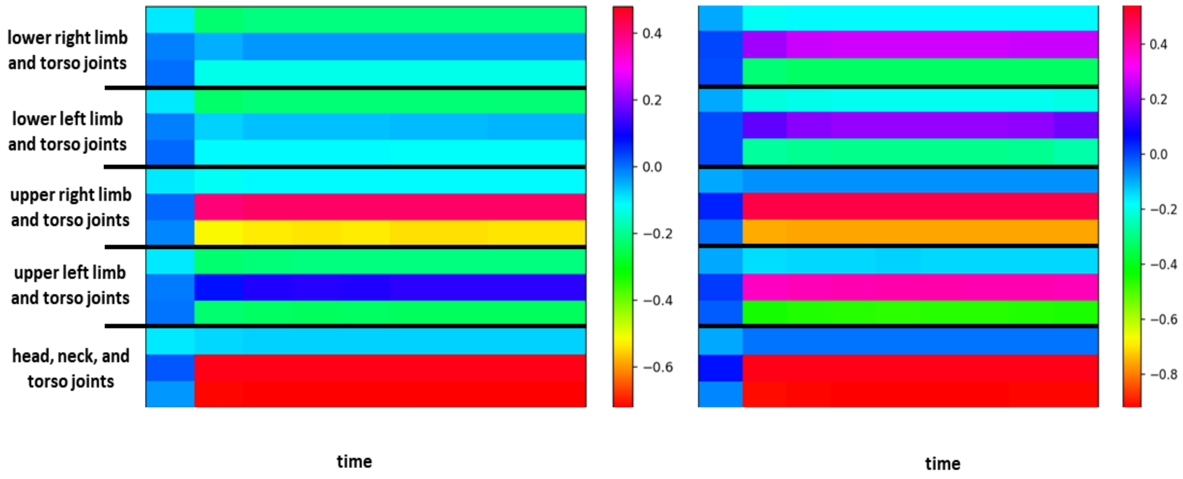


Figure 11: Time encoder activation for two different exercise segments (left, right) and two different exercise types and participants

Building on the BANet which had different but shared time encoding for different groups of anatomical segments, we propose the MiMT architecture that additionally learns multiple label timescales simultaneously. Our findings suggest that the two elements are together valuable for modelling the multiple timescales in movement data. They further highlight the importance of investigating timescales of movement assessment as is one of the aims of our observation study with physiotherapists. We plan to extend the MiMT by integrating it with other multiple timescales machine learning architectures.

## Study 3: Multimodal Movement Data Fusion based on the GWN

The GWN addresses the differences in timescales between multiple modalities of movement, using the machine learning attention (i.e. weighting) mechanism for fusion similar to the BANet although the attention module it uses is based on self-attention such that each modality assigns weights to itself and each of the other modalities. In-depth description of the GWN can be found in the peer-reviewed publication of the study.
In this study, we use both the 3D full-body positions and the muscle activity data of the exercise instances in the EmoPain dataset. Since the exercise instances were of varying lengths, zero padding at the start of each instance was used to make them of uniform

lengths. To increase the data size, i.e. the number of instances, the same mirror reflection of duplicates used in Study 2 was used here except that the reflection was only done around the y-axis. Three rotation angles were used (90°, 180°, 270°) resulting in 800 data instances in total. The labels for the instances from the healthy control participants was set to *no chronic pain*. The instances from the participants with chronic pain was labelled as *with chronic pain*. The instances from this group of participants was further labelled as *zero pain* if the participant reported pain intensity of 0 for that instance, *low level pain* if the pain intensity was otherwise ≤ 5, and *high level pain* for pain intensity > 5.

We explored the GWN in two separate but related classification tasks: recognition of chronic pain instances and pain level classification. The evaluation of the GWN in these tasks was based on leave-one-subject-out cross-validation. For each task, we compared the performance of the GWN with a baseline architecture where fusion of the multimodal data is based on simple fusion. Table 8 outlines the difference between the GWN and the baseline. In the EmoPain dataset, the two modalities were of the same sampling rate of 60Hz, the muscle activity data having been downsampled from its original 1000Hz. For the *recognition of chronic pain instances* task, both modalities were further resampled to 10Hz to manage the dimensionality of the training data. While the positional data has 3x26=78 dimensions, the muscle activity data has only 4. There were 64 units in LSTM layer which serves as attention time encoder and ordinary time encoder for the GWN and baseline architecture respectively. The Adam optimisation algorithm was used for training the models, with learning rate and batch size of 0.001 and 32 respectively, based on grid search.

Table 8: An overview of the GWN and the baseline used for comparison

| Architecture | Maps different sampling rates and/or degrees of freedom in the multiple modalities to a uniform sampling rate and dimensionality | Weighted fusion of multiple modalities (based on self-attention) | Propagation of the weightings over time |
|---|---|---|---|
| GWN | Yes | Yes | Yes |
| Simple concatenation | No | No | No |

The performance of the GWN is shown in Table 9. Both the GWN and the baseline comparison architecture perform much better than chance level classification (0.5 for the recognition of chronic pain instances, 0.33 for pain level classification), the GWN clearly outperforms the baseline architecture. A Wilcoxon signed rank test across the cross-validation folds indeed shows statistically significant difference between their performances for the recognition of chronic pain instances in particular.

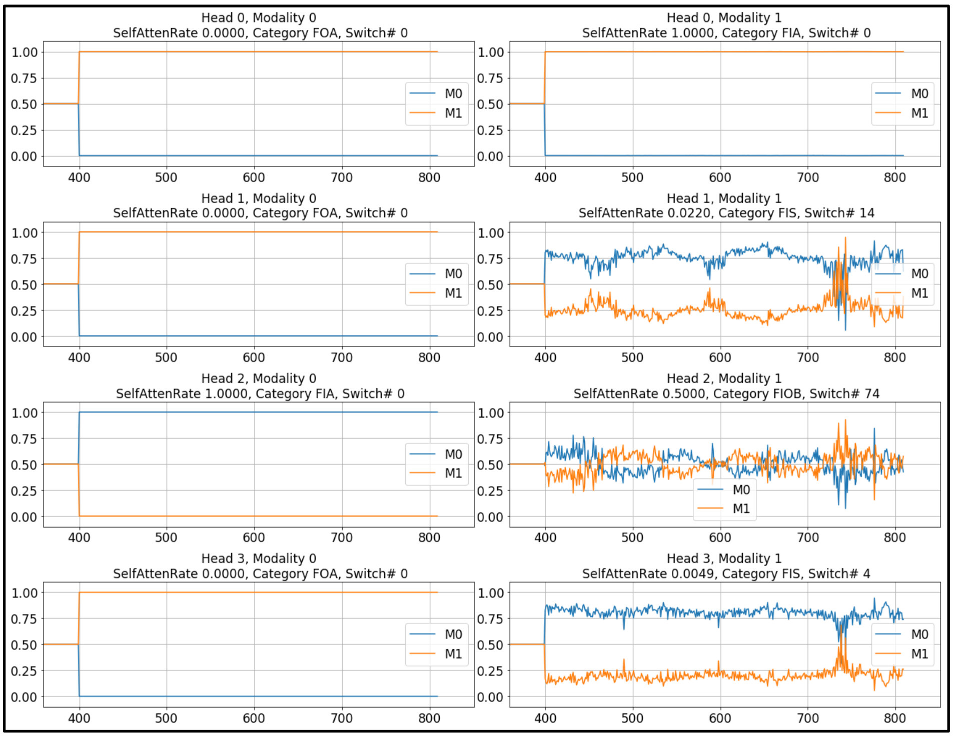
Table 9: Mean F1 scores of the GWN and the comparison architecture (in bold is the best performance and * is used to indicate performance significantly worse, p<0.05, than the GWN).

| | Mean F1 scores | |
|---|---|---|
| Architecture | Recognition of chronic pain instances | Pain level classification |
| GWN | 0.92 | 0.75 |
| Simple concatenation | 0.72* | 0.63* |

We further analysed the self-attention scores of the GWN and found 5 main temporal patterns of attention. Table 10 gives an overview of these patterns and Figure 12 gives examples of these pattern types.

Table 10: The 5 temporal patterns of self-attention found.

| Short name | Long name | Pattern (weighting are between 0 and 1 and add up to 1 by each modality) |
|---|---|---|
| FIA | Favours Itself Always | weighting for self > 0.5 100% of the time |
| FOS | Favours Other Sometimes | weighting for self < 0.5 up to 40% of the time |
| FIOB | Favours Itself and Other in Balance | weighting for self > 0.5 40-60% of the time |
| FIS | Favours Itself Sometimes | weighting for self < 0.5 less than 40% of the time |
| FOA | Favours Other Always | weighting for self > 0.5 0% of the time |



Head 0, Modality 0
SelfAttenRate 0.0000, Category FOA, Switch# 0

Head 0, Modality 1
SelfAttenRate 1.0000, Category FIA, Switch# 0

Head 1, Modality 0
SelfAttenRate 0.0000, Category FOA, Switch# 0

Head 1, Modality 1
SelfAttenRate 0.0220, Category FIS, Switch# 14

Head 2, Modality 0
SelfAttenRate 1.0000, Category FIA, Switch# 0

Head 2, Modality 1
SelfAttenRate 0.5000, Category FIOB, Switch# 74

Head 3, Modality 0
SelfAttenRate 0.0000, Category FOA, Switch# 0

Head 3, Modality 1
SelfAttenRate 0.0049, Category FIS, Switch# 4

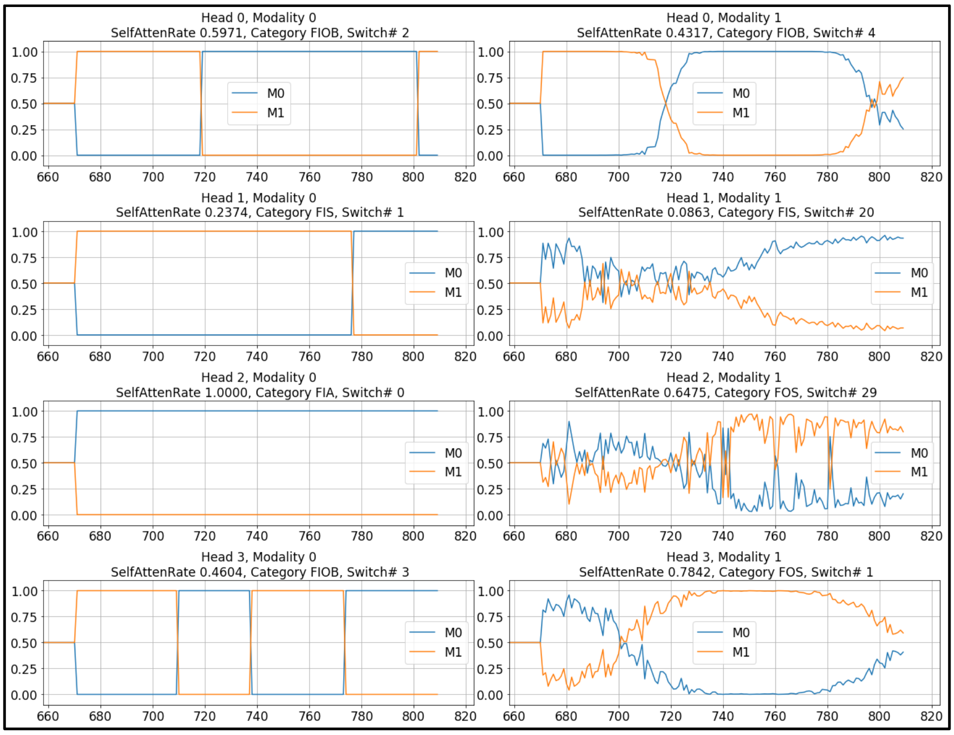


Figure 12: Plots for 2 exercise instances (top and bottom respectively) showing self-assigned attention scores versus time (M0 = positional data, M1 = muscle activity data). Plots on the left and right correspond to attention scores assigned by Modality 0 (M0) and Modality 1 (M1) respectively. The 'Head' identifier refers to the corresponding component of the attention computation ensemble; 'Switch #' refers to the number of attention switches that occur over time; and the category identifier refers to the corresponding attention pattern in Table 7.

One of the merits of the GWN approach is that it can account for noise with an unknown timescale to be accounted for. We demonstrate this by conducting an investigation of the effect of noise on the performance of the GWN and comparing the temporal patterns of self-attention with and without noise. We use Gaussian noise sampled with standard deviation equal to one-tenth of the standard deviation of the respective data modality (i.e. noise standard deviation of 10 for the positional data and 0.001 for the muscle activity data). Table 11 shows the performance of the GWN in pain level classification with and without noise in the modalities. We found no significant difference ($p$<0.05) between the performance of the GWN in both cases regardless of whether the noise was added to the positional data or to the muscle activity data suggesting that the GWN's approach to multimodal fusion indeed controls the effect of noise on the automatic detection task.

Table 11: Mean F1 scores of the GWN in the pain level classification task with and without noise in the modalities.

| | Mean F1 scores | | |
|---|---|---|---|
| Architecture | No noise | Noise in 3D position data | Noise in muscle activity data |
| GWN | 0.75 | 0.72 | 0.72 |

Table 9 shows how noise affected the distribution of the 5 temporal self-attention patterns. For the majority of data instances, the positional data modality assigns a higher weight to itself all through the time. For most of the remaining instances, this modality assigns a higher weight to the other modality all through time. The muscle activity modality also assigns a higher weight to itself all through time for a majority of the data instances, but unlike the positional data, for most of the remaining instances it instead shows the FOS pattern where it still assigns a higher weight to itself and not the other modality through most of time. Although this patterns distribution persists when noise is added to the muscle activity data, when noise is added to the positional data the distribution changes such that the positional data assigns higher weights to the muscle activity data all through time for much more data instances than those for which it assigns higher weights to itself all through time. This further highlights that the GWN enables noise in the modalities to be addressed in its fusion of multiple modalities. We speculate that the lack of difference in pattern distribution when noise was added to the muscle activity data is perhaps due to the lower dimensionality (4) of that modality, and so lower impact of noise overall, compared to that (78) of the positional data.

Table 12: The relative frequency of the 5 temporal attention patterns for each modality (M0=positional data, M1=muscle activity data). See Table 7 for the description of the patterns.

| | FIA | | FOS | | FIOB | | FIS | | FOA | |
|---|---|---|---|---|---|---|---|---|---|---|
| | M0 | M1 | M0 | M1 | M0 | M1 | M0 | M1 | M0 | M1 |
| No noise | 0.51 | 0.40 | 0.04 | 0.29 | 0.03 | 0.05 | 0.05 | 0.15 | 0.37 | 0.11 |
| Noise in M0 | 0.31 | 0.43 | 0.08 | 0.36 | 0.02 | 0.05 | 0.11 | 0.10 | 0.48 | 0.07 |
| Noise in M1 | 0.50 | 0.46 | 0.02 | 0.27 | 0.02 | 0.05 | 0.06 | 0.09 | 0.41 | 0.13 |

We propose the GWN which fuses data from multiple modalities with different sampling rates and/or dimensionalities. We showed that the GWN not only outperforms simple concatenation of these data for pain classification based on positional and muscle activity data but its good performance persists even in the presence of noise of an unknown timescale in either of the two modalities. While the modalities used in our empirical study had been (re)sampled to the same sampling rate, the timelines and timescales of events in the two modalities could still be different.

## Study 4: Accounting for Timescale Differences in Leveraging Human Activity Recognition (HAR) to enable Protective Behaviour Detection (PBD) with An Hierarchical HAR-PBD Model

The Hierarchical HAR-PBD model addresses the importance of understanding the activity being performed in a movement to determine whether or not protective behaviour is expressed in the movement while accounting for differences between the temporal structures of the activity being performed and expressions of protective behaviour. The HAR and PBD modules of the model are based on graphical convolution (Kipf and Welling 2017) and long short-term memory Hochreiter and Schmidhuber 1997; Gers et al. 1999) layers which we together refer to as GC-LSTMNN (see deliverable D3.7). In-depth description of the Hierarchical HAR-PBD model can be found in the peer-reviewed publication of the study.

The Hierarchical HAR-PBD model was evaluated using the 3D full-body joints positions from the EmoPain dataset with the corresponding continuous-time annotations of protective behaviour as well as continuous-time annotations of the activity being performed. As with Study 1, the data used was prepared by window segmentation and the protective behaviour label of each segment was defined as 1 when at least two of the four clinician raters rated at least 50% of the segment as positive for any of 5 specific protective behaviour categories (guarding, stiffness, hesitation, the use of support, and jerky motion) and 0 otherwise. The activity label of the window is based on majority vote, with six activity types in total in the data used. Further, we performed data augmentation techniques of jittering (i.e. Gaussian noise with standard deviations of 0.05 and 0.1 respectively) and cropping (i.e. drop out of timesteps and joints with selection probabilities of 0.05 and 0.1 separately) (Um et al. 2017) on these data to increase the size of the training data.

The GC-LSTMNN architecture for the HAR module was implemented as a graphical convolution network with 26 convolutional kernels, 3 LSTM (layers each with 24 hidden units, and a single fully-connected layer. A similar configuration was used for the PBD module except that 16 kernels were used in its own graphical convolution kernels. The HAR is first pre-trained and its weights frozen before being integrated in the PBD module. This strategy was used to account for possibly different temporal and/or spatial structures for HAR and PBD tasks.

To address imbalance of labels in the dataset, a novel loss function, the class-balanced focal categorical cross-entropy (Wang et al. 2021) was used. The Adam optimiser (Kingma and Ba 2014) was used for training, with learning rate of 5e-4 and 1e-3 for the HAR and PBD modules respectively. The Hierarchical HAR-PBD model was evaluated based on leave-one-subject-out cross-validation.

Table 1 shows the protective behaviour detection results of the experiments based on the Hierarchical HAR-PBD model. The performance for human activity recognition (not included in the table) was 0.88 accuracy (0.81 macro F1 score). The results show that the architecture outperforms the previous state of the art, i.e. our BANet using either angles and angular energies computed from the EmoPain dataset (Original BANet) (Wang et al. 2019) or the raw three-dimensional joints positions (Compatible BANet) in the dataset. We further see the value of including the HAR module in the architecture in the decrease in performance when the HAR module is discarded (PBD only).

In addition, we find that it was critical to pretrain this module and freeze the weights in integration for protective behaviour detection. The results indeed suggest that activity recognition only brings value to protective behaviour detection if different temporal and/or spatial representations are accounted for in the two movement abstraction levels (activity and protective behaviour).

Table 13: Results of Protective Behaviour Detection with the Hierarchical HAR-PBD Model

| Method | Accuracy | Macro F1 Score |
|---|---|---|
| Hierarchical HAR-PBD | 0.88 | 0.81 |
| Retrained-HAR Hierarchical HAR-PBD (i.e. pre-trained HAR weights not frozen) | 0.76 | 0.55 |
| Unpretrained-HAR Hierarchical HAR-PBD (i.e. HAR not pre-trained) | 0.71 | 0.45 |
| PBD only (i.e. without HAR) | 0.83 | 0.71 |
| Original BANet (Study 1) | 0.78 | 0.56 |
| Compatible BANet (Study 1) | 0.79 | 0.63 |

## References

Aung MSH, Kaltwang S, Romera-Paredes B, Martinez B, Cella M, Valstar M, Meng H, et al (2016) The Automatic Detection of Chronic Pain- Related Expression: Requirements, Challenges and a Multimodal Dataset. IEEE Transactions on Affective Computing 7(4): 1–18.

Cook KF, Keefe F, Jensen MP, Roddey TS, Callahan LF, Revicki D, Bamer AM, et al (2013) Development and Validation of a New Self-Report Measure of Pain Behaviors. Pain 154 (12): 2867–76.

Flash T, Hogans N. 1985. The Coordination of Arm Movements: An Experimentally Confirmed Mathematical Model. J Neurosci. 5:1688–1703.

Gers FA, Schmidhuber J, Cummins F (1999) Learning to forget: continual prediction with LSTM. In Proceedings of the International Conference on Artificial Neural Networks: 850–855

Hochreiter S, Schmidhuber J (1997)  Long Short-Term Memory. Neural Comput. 9(8): 1735–1780.

Keefe FJ, Block A (1982) Development of an Observation Method for Assessing Pain Behavior in Chronic Low Back Pain Patients. Behavior Therapy 13 (4): 363–75.

Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. In Proceedings of the International Conference on Learning Representations (ICLR).

Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Olugbade TA, Singh A, Bianchi-Berthouze N, Marquardt N, Aung MSH, Williams A (2019) How Can Affect Be Detected and Represented in Technological Support for Physical Rehabilitation? Transactions on Computer-Human Interaction.

Olugbade T, Newbold J, Johnson R, Volta E, Alborno P, Niewiadomski R, Dillon M, Volpe G, Bianchi-Berthouze N (2020) Automatic Detection of Reflective Thinking in Mathematical Problem Solving based on Unconstrained Bodily Exploration. IEEE Transactions on Affective Computing.

Um TT, Pfister FM, Pichler D, Endo S, Lang M, Hirche S, Fietzek U, Kulić D (2017) Data augmentation of wearable sensor data for Parkinson’s disease monitoring using convolutional neural networks. In Proceedings of the 19th ACM International Conference on Multimodal Interaction: 216-220.

Wang C, De C Williams AC, Lane ND, Bianchi-Berthouze N (2021) Leveraging Activity Recognition to Enable Protective Behavior Detection in Continuous Data. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5 (2): 1–24.

Wang C, Peng M, Olugbade TA, Lane ND, De C Williams AC, Bianchi-Berthouze N (2019) Learning temporal and bodily attention in protective movement behavior detection. In Proceedings of the International Conference on Affective Computing and Intelligent Interaction Workshops and Demos: IEEE, 324–330.

Wang C, Olugbade TA, Mathur A, De C Williams AC, Lane ND, Bianchi-Berthouze N (2019) Recurrent network based automatic detection of chronic pain protective behavior using mocap and semg data. In Proceedings of the 23rd International Symposium on Wearable Computers:225–230.

### 2.1.1 Cortico-motor alpha coherence influence visual perception

**For a full description please see:** Tomassini A., Maris E., Hilt P.M. Fadiga L., D'Ausilio A. (2020) Visual detection is locked to the internal dynamics of cortico-motor control. PLoS Biol, 18(10):e3000898.

Movements overtly sample sensory information, making sensory analysis an active-sensing process (Scott et a,. 2015; Engel et al., 2001; Palva and Palva, 2018; Tomassini et al., 2017). In this study, we show that visual information sampling is not just locked to the (overt) movement dynamics, but to the internal (covert) dynamics of cortico-motor control. We asked human participants to perform an isometric motor task – based on proprioceptive feedback – while detecting unrelated near-threshold visual stimuli. The motor output (Force) shows zero-lag coherence with brain activity (recorded via electroencephalography) in the beta-band, as previously reported. In contrast, cortical rhythms in the alpha-band systematically forerun the motor output by 200ms. Importantly, visual detection is facilitated when cortico-motor alpha (not beta) synchronization is enhanced immediately before stimulus onset, namely at the optimal phase relationship for sensorimotor communication. These findings demonstrate an ongoing coupling between visual sampling and motor control, suggesting the operation of an internal and alpha-cycling visuomotor loop.

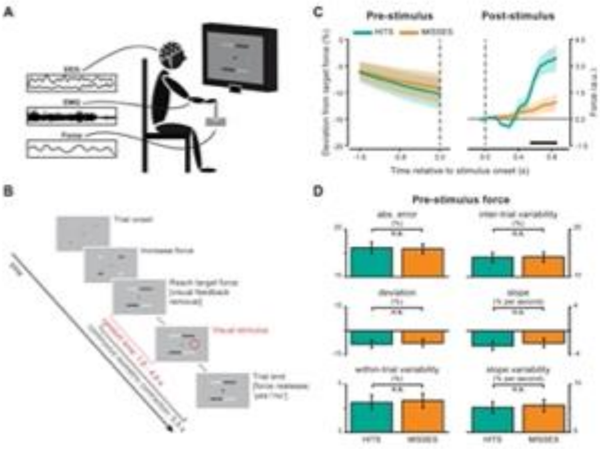
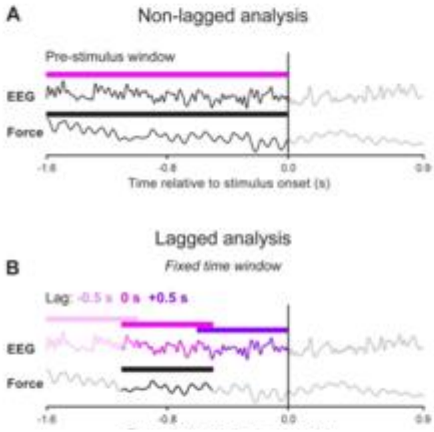


**Fig 1**. **Experimental set-up, procedure and behavioral results**. (A) EEG, EMG and Force were recorded while participants performed two tasks concurrently: visual detection and right wrist abduction to push an isometric joystick's handle towards one's own body. (B) Visual feedback of the force (four horizontal bars elongating towards the center of the screen) was provided until participants reached the target force level. Afterwards, participants were required to fixate and maintain stable contraction for 5.5 s without visual feedback. During continuous contraction, a near-threshold visual dot could appear 7.5º to the right of fixation and at a random time between 1.6 and 4.6 s (no stimulus was presented in 15% of trials). Trial end was signaled by a question mark prompting participants to release the contraction and report verbally whether they had seen or not seen the visual stimulus. (C) Force time courses in the pre- (left) and post- (right) stimulus period for hits- and misses-trials. Shaded areas represent ± 1 standard error of the mean (SEM). The black horizontal line indicates the time interval (0.55-0.85 s) belonging to the cluster that survived cluster-based permutation statistics for the hits-misses contrast. (D) Motor performance in the -1.6-0-s-window before stimulus presentation quantified as absolute (error) and relative (deviation) percentage difference from target force, inter-trial and within-trial force variability, slope and slope variability. Error bars represent ± 1 SEM.

**Fig 2**. **Schematic representation of the main analyses**. All panels show the entire available epoch [time-locked to stimulus onset: from -1.6 to 0.9 s] for example force and EEG signals. Different panels illustrate how the windowing and time-shifting (if applicable) of the signals has been performed for the main analyses. (A) Coherence and Granger causality are computed between force (black) and EEG (violet) data windows encompassing the entire pre-stimulus epoch [i.e., from -1.6 to 0 s]. (B) Lagged coherence is computed between a fixed 0.6-s force window (black) centered at -0.8 s [extending from -1.1 to -0.5 s] and 0.6-s EEG data windows that are either time-aligned with the force (violet; lag: 0 s) or shifted in time by up to 0.5 s (in 10-ms steps) in the backward (pink; lag: -0.5 s) and forward (dark violet; lag: +0.5 s) direction. (C) Time-resolved lagged coherence is computed between 0.3-s force windows that are advanced over time (in 10-ms steps) from -1 s (black) up to a variable time point depending on the analysis (gray; example time is 0 s for illustrative purposes) and corresponding 0.3-s EEG data windows that are either time-aligned with the



force (violet; lag: 0 s) or shifted in time by up to 0.4 s (in 10-ms steps) in the backward (pink; example lag is -0.2 s for illustrative purposes) and 0.2 s in the forward (dark violet; lag: +0.2 s) direction.

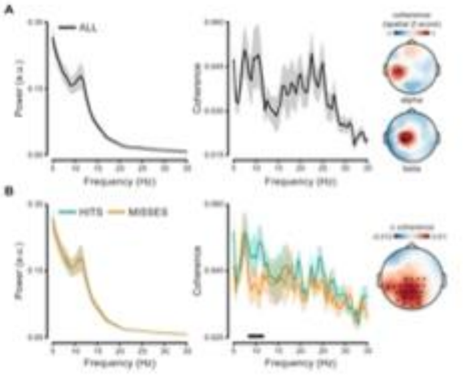


**Fig 3**. **Spectral content of pre-stimulus force and coherence with cortical activity**. (A) Force power (left) and coherence (middle) spectrum with contralateral centro-parietal EEG electrodes (C1, C3, C5, CP5; marked in grey in the topographic maps) computed on the pre-stimulus window (-1.6-0 s). Topographies show coherence in the alpha (9-11 Hz; top) and beta (20-30 Hz; bottom) range. Coherence has been spatially z-scored before averaging across subjects and frequencies by subtracting the individual mean coherence over the electrodes and dividing the result by the standard deviation across the electrodes. (B) Same as in (A) but computed separately for hits- and misses-trials (left, middle). The black horizontal line indicates the frequency interval (8.5-11.5 Hz) belonging to the cluster that survived cluster-based permutation statistics for the hits-misses contrast. Coherence spectra are averaged over the EEG electrodes belonging to the same cluster (evaluated at 10.5 Hz; see black asterisks in the topographic map). Topography shows the hits-misses difference in coherence averaged over the cluster frequency interval (8.5-11.5 Hz).

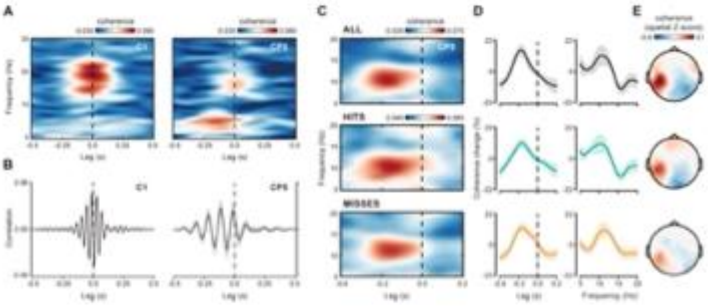


**Fig 4**. **Lag-dependency of cortico-force coherence**. (A) Lag- and frequency-resolved cortico-force coherence is shown for two EEG electrodes, C1 (left) and CP5 (right), where beta- and alpha-band coherence is maximal, respectively. Coherence has been calculated on 0.6-s data windows (from -1.1 to -0.5 s) by shifting the EEG signal (relative to the force signal) by a variable amount of time (negative lags: EEG precedes force; positive lags: EEG follows force)  (B) Cross-correlation between force

and EEG activity [over the same electrodes as in (A)] that was previously band-pass filtered (zero-phase filtering by two-pass Butterworth, 2nd order) in the beta (20-30 Hz; left) and alpha (8-12 Hz; right) range. Cross-correlations are normalized so that the autocorrelations at zero lag are identically 1. (C) Lag-frequency coherence representation as in (A) but computed on shorter (0.3-s) sliding data windows and then averaged over the pre-stimulus period for all trials as well as separately for hits- and misses-trials. (D) Lag (left) and spectral (right) tuning of cortico-force coherence expressed as the relative percentage change in coherence averaged over frequencies between 8 and 12 Hz and lags between -0.36 and 0 s (i.e., lag of max. alpha coherence on all trials [-0.18 s] ±1SD across subjects), respectively. (E) Topographies show coherence at frequency 10.5 Hz and lag -0.2 s for all trials (top), hits (middle) and misses (bottom).

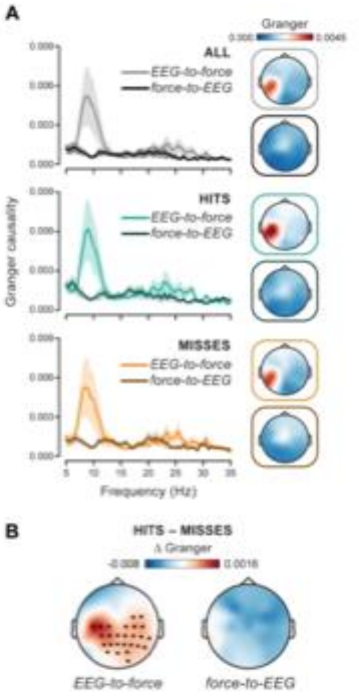


**Fig 5**. **Cortical alpha drives alpha fluctuations in the force: Granger causality**. (A) Granger causality in the EEG-to-force and force-to-EEG directions (evaluated at electrode CP5) computed on the entire pre-stimulus interval (-1.6 – 0 s) for all trials (top), hits (middle) and misses (bottom). Topographies show Granger causality in both directions (top: EEG-to-force; bottom: force-to-EEG). (B) Topographies show the hits-misses difference in Granger causality (left: EEG-to-force; right: force-to-EEG) evaluated at frequency 10.5 Hz (black asterisks mark electrodes belonging to the cluster that survived cluster-based permutation statistics).

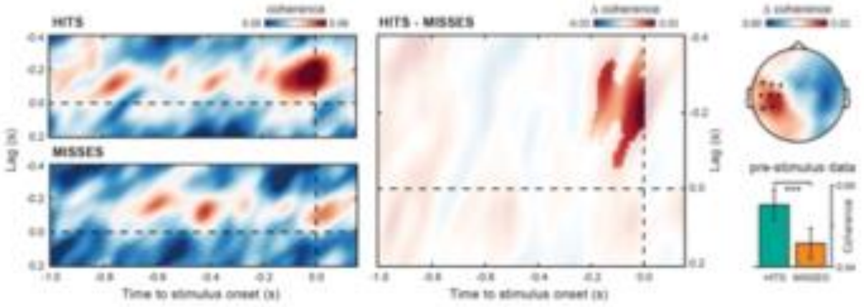


**Fig 6**. **Cortico-force alpha coherence just before stimulus onset predicts perception**. Lag- and time-resolved cortico-force alpha (10.5 Hz) coherence over the pre-stimulus period for hits, misses and their difference (hits-misses). The highlighted area indicates the time and lag intervals belonging to the cluster that survived cluster-based permutation statistics for the hits-misses contrast. Alpha coherence is averaged over the EEG electrodes belonging to the same cluster (evaluated at time 0 s and lag -0.2 s; see black

asterisks in the topographic map). The topography shows the hits-misses difference in alpha coherence averaged over the time and lag intervals belonging to the same cluster. The bar plot shows alpha coherence for the electrode CP5, calculated at lag -0.2 s and time -0.16 s, i.e., the time point closest to stimulus onset where the analyzed data windows do not include any post-stimulus data point. Error bars indicate ± 1 SEM. ****p*<0.001.

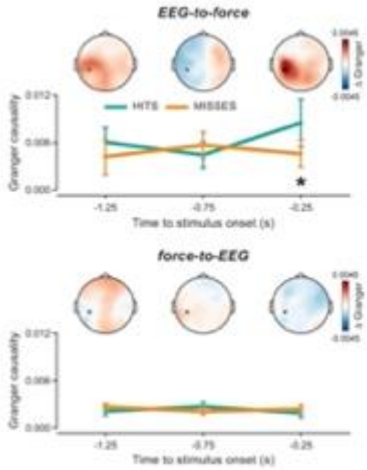


**Fig 7**. **Time-resolved Granger causality: EEG-to-force alpha connectivity just before stimulus onset predicts perception**. Granger causality in the EEG-to-force (top) and force-to-EEG (bottom) directions evaluated at frequency 10.5 Hz and electrode CP5 (marked in gray in the topographic maps) is shown for hits and misses as a function of time before stimulus onset (i.e., for three non-overlapping pre-stimulus 0.5-s time windows centered at -1.25, -0.75 and -0.25 s). Topographies show the hits-misses difference in Granger causality at corresponding times (see above; the black asterisk indicates that electrode CP5 survived permutation statistics for the hits-misses contrast; p = 0.0158).

## References

Engel, A.K., P. Fries, and W. Singer, Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci, 2001. 2(10): p. 704-16.

Palva, S. and J.M. Palva, Roles of Brain Criticality and Multiscale Oscillations in Temporal Predictions for Sensorimotor Processing. Trends Neurosci, 2018. 41(10): p. 729-743.

Scott, S.H., et al., Feedback control during voluntary motor actions. Curr Opin Neurobiol, 2015. 33: p. 85-94.

Tomassini, A., et al., Theta oscillations locked to intended actions rhythmically modulate perception. Elife, 2017. 6.

## 2.1.17 Intersecting action and perception in autism spectrum disorders at the single-trial level

*Noemi Montobbio, Andrea Cavallo, Dalila Albergo, Caterina Ansuini, Francesca Battaglia, Jessica Podda, Lino Nobili, Stefano Panzeri, Cristina Becchio (in preparation).*

Individuals with autism spectrum disorders (ASD) struggle to attribute intention to action. However, the computational and neurobiological bases of these difficulties remain poorly understood. In this section, we report on the results of a study combining motion tracking, psychophysics, and computational analyses to uncover intention readout computations in typically developing (TD) children and children with ASD watching actions produced by typical and autistic children.

Eight- to thirteen-year-old TD children (N = 35) and children with ASD (N = 35) watched a hand reaching for a bottle and judged on the intention of the observed grasp. To capture natural movement variability, we selected 100 representative reach-to-grasp actions (50 ASD actions and 50 TD actions) from a large action dataset obtained by filming and simultaneously tracking TD and ASD children reaching toward and grasping a bottle with the intent to place or pour. In a within-subjects counterbalanced order, participants watched videos of actions performed by TD children and ASD children (Figure 1A-C).

We developed a novel framework to quantify how intention encoding (how intention information is encoded in movement kinematics) and readout (how intention information is readout from visual kinematics) intersect at the single-trial level in typical and autistic observers (Panzeri et al., 2017; Patri et al., 2020). We first quantified intention information in TD and ASD single-movement kinematics and determined the set of kinematic features that encode such information. We then developed a readout model to quantify how (and how well) typical and autistic observers read intention information in typical and autistic visual kinematics. We finally examined how encoding and readout intersect at the single-trial level. This approach allowed us to move beyond representations averaged over trials and participants and determine how individual observers read out intention information at a single-trial level to inform intention choices.

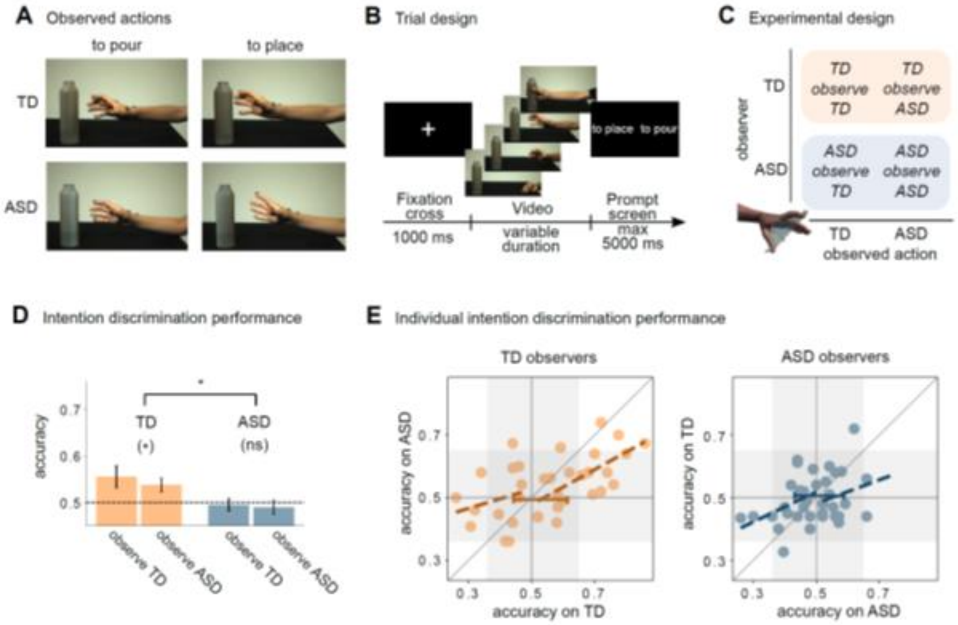


**Figure 1. Experimental design and intention discrimination results.** (**A**) Example video frames of grasp-to-pour and grasp-to-place acts produced by TD and ASD children. Each video started at reach onset and ended at the contact time between the hand and the bottle. (**B**) Trial design of the intention discrimination task. (**C**) Schematic of the experimental design. (**D**) Trial-averaged intention discrimination performance (fraction correct) for each observer group and observed action. (**E**) Scatter plot of individual intention discrimination performance on TD actions against ASD actions. For each observer group, regression lines estimated via piecewise regression analysis and 95% confidence intervals of the breaking point are displayed.

Trial-averaged intention discrimination was above in TD children, but not in ASD children (Figure 1D). However, single-trial analysis revealed that both TD and ASD intention choices reflected systematically trial-to-trial variations in visual kinematics (Figure 2).

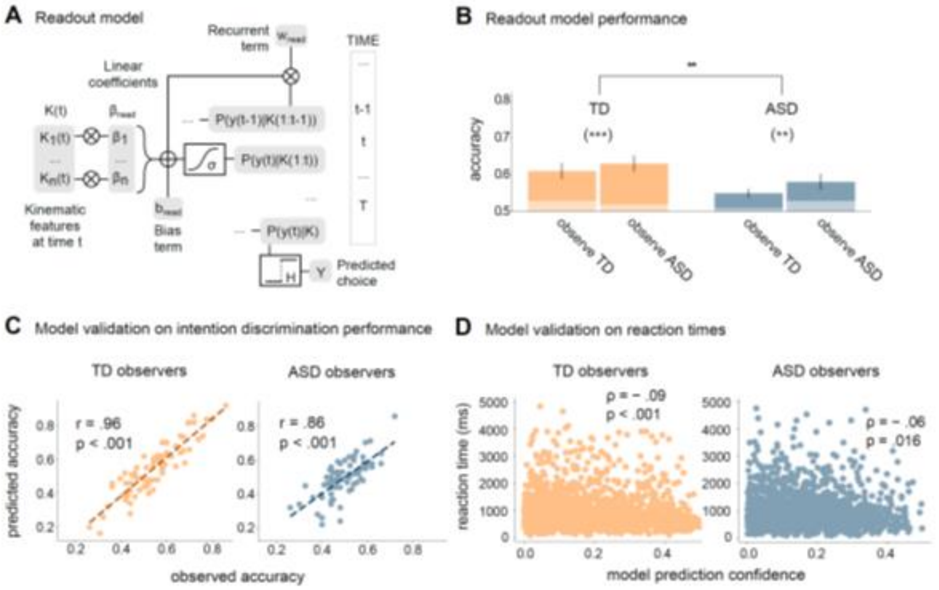


**Figure 2. Readout of intention from single-trial kinematics**. (**A**) Block-diagram and equation of the model used to quantify intention readout. (**B**) Cross-validated performance of the readout models quantified as fraction of correctly predicted intention choices. Light sub-bars represent the chance level performance quantified as the mean of the null-hypothesis distribution of cross-validated model performance. (**C**) Scatter plots of the relationship between the observed intention discrimination performance and the one predicted by the readout model across individual participants, for TD and ASD observers separately. Pearson's correlation coefficients and their significance values are reported. (**D**) Scatter plots of the relationship between trial-level reaction times and model prediction confidence, computed as the deviation of the estimated probability of 'to pour' from chance. Spearman's correlation coefficients and their significance values are reported.

Thus, both TD and ASD observers read single-trial variations in movement kinematics, but in different ways. TD readers were better able to identify intention-informative features during observation of TD actions; conversely, ASD readers were better able to identify intention-informative features during observation of ASD actions, suggesting a kinematic similarity advantage. Crucially, while TD observers were generally able to correctly interpret the extracted intention information, those with autism were unable to do so, regardless of whether the information was extracted from TD or ASD visual kinematics.

These results expand existing conceptions of intention reading in autism by suggesting that difficulties in attributing intention to action in autism are specifically the result of a deficit in linking informative kinematics to intention.

## References

Panzeri, S., Harvey, C. D., Piasini, E., Latham, P. E., & Fellin, T. (2017). Cracking the Neural Code for Sensory Perception by Combining Statistics, Intervention, and Behavior. Neuron, 93(3), 491-507. https://doi.org/10.1016/j.neuron.2016.12.036

Patri, J.-F., Cavallo, A., Pullar, K., Soriano, M., Valente, M., Koul, A., Avenanti, A., Panzeri, S., & Becchio, C. (2020). Transient Disruption of the Inferior Parietal Lobule Impairs the Ability to Attribute Intention to Action. Current Biology, 30, 4594-4605. https://doi.org/10.1016/j.cub.2020.08.104