DISSEMINATION LEVEL: PU

D2.5
Results on prediction in complex action execution and observation - Phase I

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

Table of Contents

Introduction

This deliverable reports on the progress on the research conducted between M7-M24 of the EnTimeMent project with regards to the complex action execution and observation axis, focused on studies on group motor behaviour (involving three people or more, n>2). This part of the project is fed by the research and theoretical work developed in the Phase I of the EnTimeMent project, focused on the individual (n=1) and dyadic studies (n=2), which are reported in D2.1 and D2.3 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 *(2.3.1; 2.3.2; 2.3.5).* For *2.3.3 Time to sync,* the data collection is ongoing for two out of four experiments planned, with three experiments being granted ethical approval from a local IRB committee (please see in the Appendices IRB 2002A, 2002B and 2005A) at EuroMov. The data collection for *2.3.4 Similarity of motor signature across multiple timescales in musical performers* is about to start at UNIGE.

Our brains have been carved by evolution to act together with other people, towards long-term mutual goals. Studies reported below investigate how people combine sensory information to achieve coordination with others, drawing in parallel from their internal mental representations (i.e. memory), adaptation and prediction processes. First project, *2.3.1* at IIT-FE, harnessed a naturalistic interaction in the violin ensemble to understand those sensorimotor dependencies. Two channels of communication were explored (head and bow movements of the violin players - section one and two; in relation to the conductor) under full and perturbed access to the visual information of other players. Study has found that bow kinematics in players in two violinists' sections exhibit a robust leader-follower relationship with the conductor. Study showed that bow kinematics resist visual perturbation (part of the section being deprived of the vision of the other part) due to reliance on memory of the score and aforementioned primary role of flow of information from conductor in leading the group performance. For the head movement of players, different pattern emerged revealing emergence of strategy to maximize performance accuracy in the section deprived of visual contact with conductor by tighter coordination of within the section (see Figure 1 and 2). In parallel, drop in intra-group coordination was noted for the section with preserved visual contact with the conductor, but becoming a coordination hub for the section deprived of visual information and conductor. This emergence of strategy demonstrated the ability of experts, such as violin players, to adapt intra- and inter-group dynamics to maintain robust coordination in a context of a group perceptuomotor performance. Next project, *2.3.2* continued to further explore how humans (expert- dancers; and non-expert - sport students) maintain regimes of coordination despite a transient loss of visual coupling during pendulum-based synchronization task. This study has strengthened the idea that the ‘topology’ - spatial alignment between people mediates access to the real-time, perceptual information about movement of others plays a key role in the succeeding in behavioural cohesion. Participants were able to maintain synchronization of swinging the pendulum with matching frequencies up to 7 seconds after the loss of perceptual contact (manipulation of field of view with visual occlusion glasses), with more accurate performance observed for experts, revealing both short timescale memory constraints and role of perceptuo-motor expertise (long timescale) in achieving coordination success. Finally, the role of the perceptual contact intra-group and with the leader, was dissected in the *2.3.5* project looking into the eye contact (gaze) relationship during group discussion with manipulation of the factors of the leadership style (authoritarian versus democractic, and high and low time pressure condition). Leaders looked less at others and, conversely, were looked at more as compared with followers. Also, leaders were involved in and caused more episodes of mutual engagement, relative to followers. This study has demonstrated the key contribution of the gaze exchange in social interactions, building up on the previous results from 2.3.1 and 2.3.2. Taken together, Phase I results reported herein pushed forward current state-of-the-art models of human coordination in complex group situations involving both action execution and observation. Studies reported below addressed multiple gaps in the body of research and emphasised importance of multiple timescales approach in analysis of human sensorimotor group behaviour. Research roadmap for Phase II of EnTimeMent project has been established (D1.2 Research requirements), which will push further the frontiers towards a full understanding of the sensorimotor dependencies of human interaction.

2.3.1 Orchestra violin sections and conductor

For a full description please see: Hilt P.M., Badino L., D'Ausilio A., Volpe G., Fadiga L., Camurri A. (2019) Multi-layer adaptation of group coordination in musical ensembles. Sci Rep, 9: 5854.

Successful human-to-human interaction requires important behavioral adaptation, as well as prediction. A large body of literature has focused on cooperation towards shared goals, where humans must combine available sensory information with internal movement production models (Wolpert et al., 2013; Sebanz et al., 2009; Jeannerod et al. 2001; Friston et al. 2011). In this regard, researchers investigated how dyads achieve interpersonal simple sensorimotor coordination, such as walking side-by-side (van Ulzen et al, 2015) or rocking in rocking-chairs (Richardson et al. 2007). In such contexts, co-actors continuously influence each other and tend to spatially and temporally synchronize their movements. Beside imitation, action complementarity play a key role in inter-individual coordination with the goal of achieving efficient collaboration (Newman-Norlund et al., 2007). Social interaction indeed goes beyond synchronization with other's actions and relies also on inferring others' motor goals and intentions to generate a context-appropriate action. To achieve fast inter-individual coordination, individuals may build internal predictive models of other's behavior. In function of the context, the most appropriate motor model is compared with the current observed movement, to generate a prediction error (Friston et al., 2011) and update own motor planning (Sebanz et al., 2006). Due to the technical and analytical complexity in exploring the details of human sensorimotor interaction, only few experiments went further than a dyadic set-up (Fessler et al., 2016; Dikker et al., 2017, Alderiso et al., 2016). However, in daily life, things are usually much more complex. For instance, during a conversation, information is sampled through multiple channels (e.g. vision, audition), sometimes in parallel (e.g. information in the foreground and information from the background) and at different temporo-spatial scales (e.g. slow whole-body movements versus fast lip motions). At the same time, different kinds of information may be conveyed in parallel through different channels. For example, in speech, bodily gestures and spoken words are generally co-expressive (McNeill, 2000). On this basis, communication requires flexible means to integrate multimodal data, across multiple timescales and act accordingly. Therefore, proper quantification of (realistic) group coordination is today one of the key missing elements to understand how humans manage to interact with others by efficiently selecting, processing and sending information.

In this context, ensemble musicians have been proposed as an ideal model, by keeping the key multidimensional properties of natural sensorimotor interaction, but allowing relatively good experimental control (Volpe et al., 2016; D'Ausilio et al. 2015). Beyond global descriptions of musician's pattern of relationships, the complexity of these kinds of scenarios could also be exploited to distinguish and evaluate the existence of multiple channels of communication as well as their respective role in efficient coordination. In previous studies, one representative kinematic parameter was used to extract global coordination (D'Ausilio et al., 2012; Badino et al., 2014, Chang et al., 2017., Chang et al., 2019). However, we know that movements of different body parts may convey substantially different types of information. For instance, bow movements in violinists directly control the sound output (i.e., instrumental gestures), whereas complementary torso oscillations may serve a secondary communicative purpose (ancillary gestures (D'Ausilio et al., 2015)). More importantly, movements of different body parts may act as different channels of communication, possibly with different roles depending on the specific communication mode. For example, within a quartet (Badino et al., 2014; Chang et al., 2017; 2019), musicians have specific roles while in orchestras, musicians generally play in distinct sections (e.g. sections of violinists). This means that in the orchestra scenario, different modes of communication coexist: a complementary coordination with the conductor and other musicians, in parallel with an imitative coordination with musicians of the same group (playing the same score).

In the present study, we aim at answering two scientific questions: whether different channels of communication exist and whether they carry different information across modes of communication. We had a chamber orchestra playing music while we recorded bow and head kinematics (instrumental and ancillary movements) of a first and second section of violinists (four violinists in each section) as well as the arm and head kinematics of two different conductors. In one experimental condition we applied a perturbation to the orchestra sensorimotor information flow. The perturbation consisted in half-turn rotation of the first section of violinists so that they faced the second section and couldn't see the conductor anymore. This perturbation modifies the perceptuo-motor context of the first section of violinists, placing also the second section and the conductor into a novel playing situation. By doing so, we analyzed inter-group complementary coordination as well as intra-group imitative coordination (modes of communication), through different channels of communication (instrumental and ancillary movements) during different playing situations (normal and perturbed).

Our results demonstrate that the pattern of sensorimotor information carried by two selected movements (head and bow) are distinct. Bow kinematics exhibit a robust leader-follower relationship between the conductor and the two violinists' sections. This pattern is substantially not affected by the experimental manipulation of the sensorimotor information flow (perturbed condition) except for a decrease in communication between the first section and the conductor. The fact that the perturbation did not dramatically alter the information exchanged via instrumental movements suggests an important role of memory, score reading and residual sensory cues. Indeed, musicians train for several hours and may rely on rehearsal memory to cope with the perturbation, at least for what concerns pure instrumental execution. At the same time, there is also a clear directionality of the information flow from conductor to musicians, which confirms the idea of a predominant role of the conductor in group management. Ancillary movements, instead, are supposed to convey slower frequency signals possibly related to the expressive component of musical execution, which is more likely to be affected by perturbation of the interaction dynamics. In fact, in head data, the perturbation produced clear alteration of the communication pattern. Communication between the first section and the conductor or the second section was reduced. At the same time, communication between the second section and the conductor increased in both directions. This global increase suggests a greater need for information exchange during the perturbation and considering that conductors and S2 did not change their positions, we observe a quantitative but not a qualitative alteration of their communication. Instead, moving to the relationship between S1 and S2 we observe a complete reversal of their mutual communication. Before the perturbation, the first section provided larger causal drive towards the second section, while after, the second section led the first. During the perturbation, the first section no longer had visual contact with the conductor, significantly reducing his role in leading orchestra dynamics. Even if we cannot exclude the contribution of haptic, acoustic or residual visual information, the remaining influence that the conductor exert on S1, seems to be mediated by the new role played by S2. Although violinists of the second section did not actually change their position, they are the only ones establishing direct face-to-face communication with both the first section and conductor. Interestingly, they seem to increase their normal communication with conductors, while at the same time they dramatically change the way they communicate with S1. Correspondingly, our results suggest that S2 musicians were implicitly invested with far more centrality in orchestra coordination dynamics.

The distinct modulation of head versus bow kinematic parameters provides a demonstration of the multi-level complexity of musicians' coordination. At the same time, another important aspect is related to the co-regulation of different modes of interaction. In our experimental context, each violinist must exchange information with other musicians of the same section (playing the same musical score – intra-group imitative coordination – Figure 1) and with other participants (playing different parts – inter-group complementary coordination – Figure 2). We used PCA to complement inter-group Gca analysis with an estimation of intra-section imitative coordination. Both kinematic parameters highlighted similar patterns of results. Due to the lack of communication with the conductor, the first section became more coordinated, in the probable attempt to maximize performance accuracy. On the contrary, the second section that was endowed with the central role of being the communication hub, reduced intra-group coordination. This may be driven by a need to gain the necessary degrees of freedom to lead communication with S1 and be the sole interlocutor of the conductor. Therefore, here we show that to modulate inter-group dynamics, S2 violinists had to penalize imitative coordination at the intra-group level.

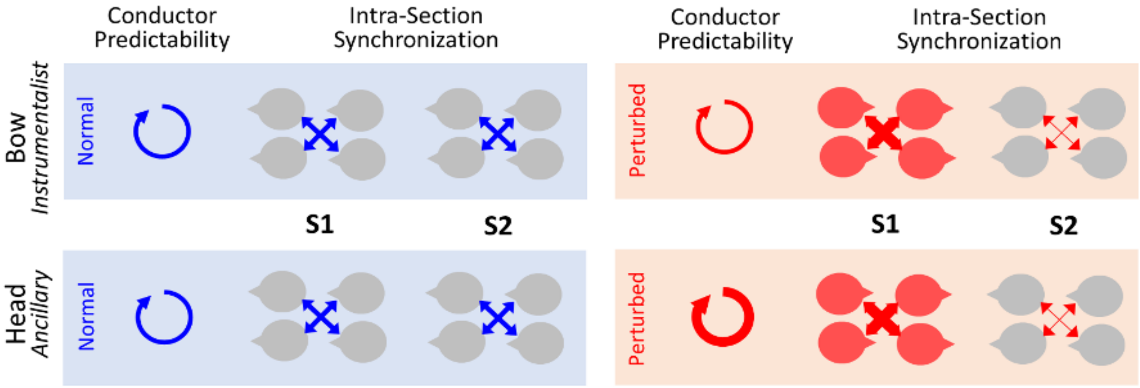


Figure 1: Schematic representation of the main results for intra-group analysis. Results associated with the two channels, bow and head, are displayed respectively in the upper and lower panel. Circular arrows displayed in the left panel represent the strength of conductor predictability (ARfit). Middle and right panels represent the intra-group synchronization (%PC1) for both sections of violinists. Thickness of the arrow represents the strength of the effect in each experimental condition: normal (blue) and perturbed (red).

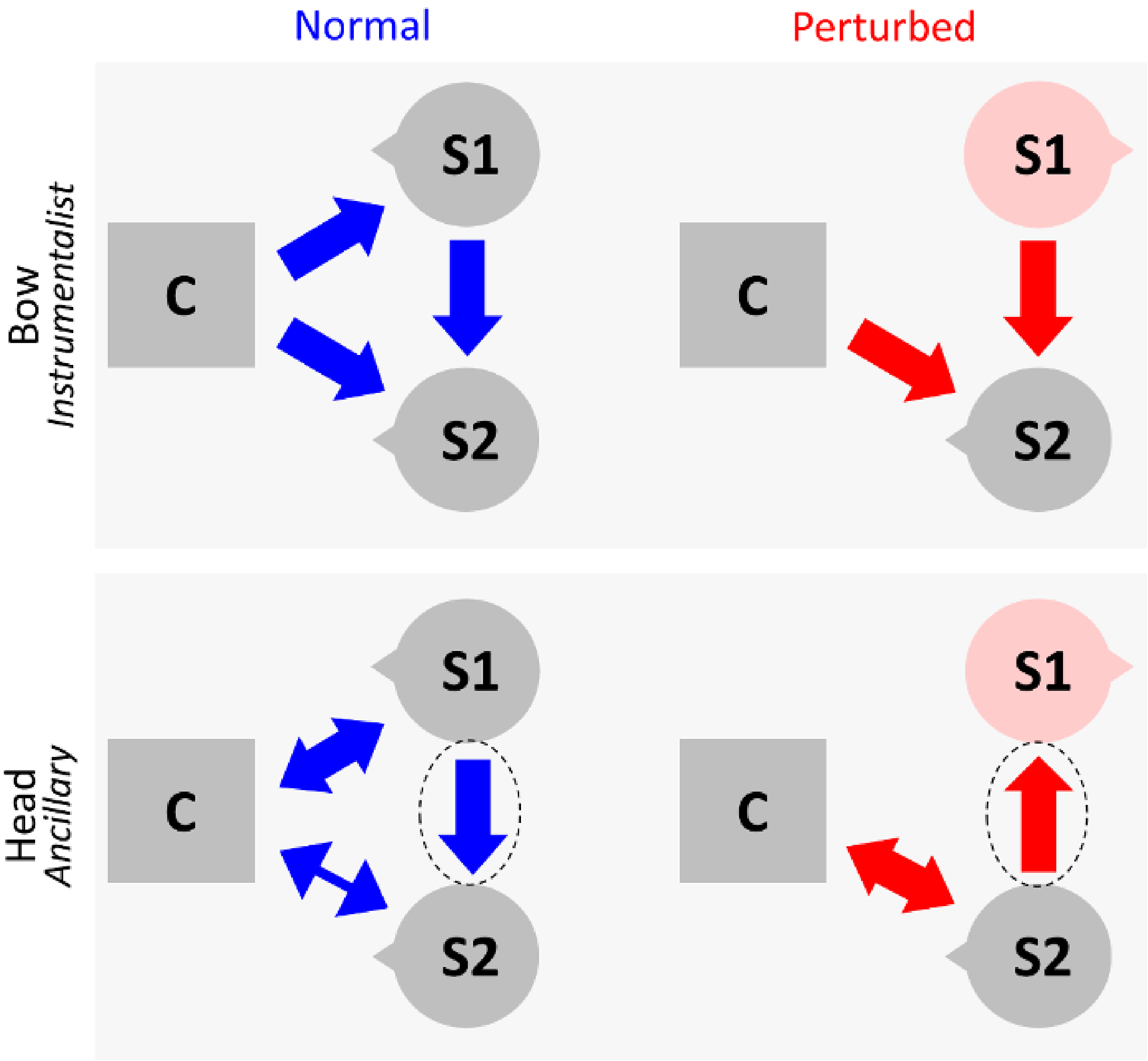


Figure 2: Schematic representation of the main results for inter-group Granger-Causality analysis (i.e. inter-group coordination) across the conductor (C) and the two sections of violinists (S1 and S2). Results associated with the two channels, bow and head, are displayed respectively in the upper and lower panel. Directional arrows illustrate inter-group coordination (C, S1 and S2), in the normal (blue) and perturbed (red) condition. Arrows thickness represents the interaction's strength. A bidirectional arrow indicates similar gca values for the two directions (i.e. group 1 G-causes group 2, as much as group 2 G-causes group 1). On the opposite, a unidirectional arrow indicates the direction of the larger gca value (e.g. group 1 G-causes more group 2, than the inverse). To highlight the difference between the two conditions, we did not represent the arrow between C and S1.

References

Alderisio, F., Fiore, G., Salesse, R. N., Bardy, B. G. & di Bernardo, M. (2016) Interaction patterns and individual dynamics shape the way we move in synchrony. arXiv. at <http://arxiv.org/abs/1607.02175>

Badino, L., D’Ausilio, A., Glowinski, D., Camurri, A. & Fadiga, L. (2014). Sensorimotor communication in professional quartets. Neuropsychologia 55, 98–104.

Chang, A., Livingstone, S. R., Bosnyak, D. J. & Trainor, L. J. (2017) Body sway reflects leadership in joint music performance. Proc. Natl. Acad. Sci. 114, E4134–E4141.

Chang, A., Kragness, H. E., Livingstone, S. R., Bosnyak, D. J. & Trainor, L. J. (2019) Body sway reflects joint emotional expression in music ensemble performance. Sci. Rep. 9.

Colley, I. D., Varlet, M., MacRitchie, J. & Keller, P. E. (2018)The influence of visual cues on temporal anticipation and movement synchronization with musical sequences. Acta Psychol. (Amst). 191, 190–200.

Codrons, E., Bernardi, N. F., Vandoni, M. & Bernardi, L. (2014) Spontaneous group synchronization of movements and respiratory rhythms. PLoS One 9.

D’Ausilio, A., Novembre, G., Fadiga, L. & Keller, P. E. (2015). What can music tell us about social interaction ? Trends Cogn. Sci. 19, 1–4.

D’Ausilio, A. et al. (2012). Leadership in orchestra emerges from the causal relationships of movement kinematics. PLoS One 7.

Dikker, S. et al. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. Curr. Biol. 27, 1375–1380.

Fessler, D. M. T. & Holbrook, C. (2016). Synchronized behavior increases assessments of the formidability and cohesion of coalitions. Evol. Hum. Behav. 37, 502–509.

Friston, K. J., Mattout, J. & Kilner, J. M. (2011). Action understanding and active inference. Biol. Cybern. 104, 137–160.

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. Neuroimage 14, S103–S109.

McNeill, D. Language and Gesture. (2000).
Newman-Norlund, R. D., Noordzij, M. L., Meulenbroek, R. & Bekkering, H. (2007). Exploring the brain basis of joint action: Co-ordination of actions, goals and intentions. Soc. Neurosci. 2.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L. & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. Hum. Mov. Sci. 26, 867–891.

Sebanz, N., Bekkering, H. & Knoblich, G. (2006). Joint action: Bodies and minds moving together. Trends Cogn. Sci. 10, 70–76.

Sebanz, N. & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where. Top. Cogn. Sci. 1, 353–367.

van Ulzen, N. R., Lamoth, C. J. C., Daffertshofer, A., Semin, G. R. & Beek, P. J. (2008). Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. Neurosci. Lett. 432, 88–93.
Volpe, G., D'Ausilio, A., Badino, L., Camurri, A. & Fadiga, L. (2016). Measuring social interaction in music ensembles. Philos. Trans. B 371, 20150377.

Wolpert, D. M., Doya, K. & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. Philos. Trans. R. Soc. London 358, 593–602.

2.3.2 Dance to Sync

For a full description please see: Bardy, B. G., Calabrese, C., de Lellis, P., Bourgeaud, S., Colomer, C., Pla, S., & di Bernardo, M. (2020). Moving in unison after perceptual interruption. *Scientific Reports, i* 10:18032 doi.org/10.1038/s41598-020-74914-z.

Humans and other animals often cooperate in small or large ensembles, for anti-predation, for producing a collective performance, or sometimes just for entertainment. Among all sorts of cooperative behaviours, synchronization in space and/or in time of the members of the group is particularly present in the human repertoire. It is often rooted in perceptuo-motor synergies in which proximal (e.g., postures, breaths) or distal (e.g., gazes, voices, hands, legs) parts of the body are delicately locked, for brief or long periods of time, in frequency and in phase (Kelso et al., 1986; Strogatz et al., 2004). In sport, music and dance moving in unison is either the goal or clearly contributes to it, and results from both (i) personalized characteristics and (ii) the way individuals are coupled together. Synchronization also requires a coupling function between the various systems involved, whether these systems are of physical, biological, or social origin (Damm et al., 2019). Perceptual contact is the most natural form of coupling between agents in a group.

Of interest for the present research is the recent discovery that certain topologies of the spatial organization of members in the group affect the strength and symmetry of perceptual coupling. The current study targeted the powerful capacity of humans to maintain regimes of synchronization despite a transient loss of perceptual (i.e., visual) coupling as it often occurs in daily scenarios. This phenomenon occurs for instance when a group of people continue to walk at the same pace even after they separate, or when dancers in a choreographic performance maintain body synchronization during a transient lack of visual connection. This capacity is a solid contributor to a wide range of social performances, in sport or at work. It relies on our practical ability to internalise previously-produced movement patterns in a social context, and to maintain them when alone for a certain amount of time. The aetiology of this ability is somewhat dual (physical and neural). One approach, the individual memory approach, considers this persistence effect as a witness of our capacity to prolong a movement pattern previously produced under a certain goal (intentional group synchronization) into a new context (solo action). In contrast, the social memory approach (Oulier et al., 2018; Nordham et al., 2018) suggests that persistence after visual interruption is the consequence of the mental simulation of the social interaction previously created. It also predicts that a certain proximity in individual movement frequencies fastens the route toward synchronization and helps to maintain a stronger coordinate regime during perceptual coupling and temporarily after its interruption but anchors these evidences into the social benefits of synchronization. Here we investigated the dynamics of voluntary synchronization, in groups composed of seven participants, manipulating their similarity, spatial organization and the presence or duration of visual coupling. Participants were engaged in an intentional group synchronization task and had to swing a pendulum in order to achieve unison in space and in time (phase synchronization).

This task was selected as (i) it is extremely easy to learn and perform, (ii) it has been documented before in a dyadic context (Schmidt et al., 1990), and (iii) it allows a simple yet precise control over each participant's natural frequency. Each trial started with an eyes-closed period, in which each player oscillated their own pendulum at their preferred pace. This was followed by an eyes-open period, where they had to reach synchronization as fast as possible. The last period was again an eyes-closed sequence, split into two-time intervals of equal length, in order to better identify the possible presence and duration of a memory effect.

We have conducted two experiments in this study. In Experiment 1, participants' similarity was controlled by manipulating the pendulum's inertia and hence the natural frequency of the players' oscillatory motion. This enabled us to evaluate the influence of the players' similarity and graph structure on the emergence and quality of group synchronization. Specifically, four conditions were considered, involving (i) individual oscillations (solo), and three collective oscillations (ii) at the same shared frequency (all matched), (iii) at the same frequency for six out of the seven players (all matched but one), and (iv) at seven different frequencies corresponding to each player's preferred pace (natural). Each group performed the experiments in four different interaction patterns among players (i.e., topologies), implemented through the combination of the spatial location of each participant and the use of home-made goggles limiting the field of vision to the desired location.

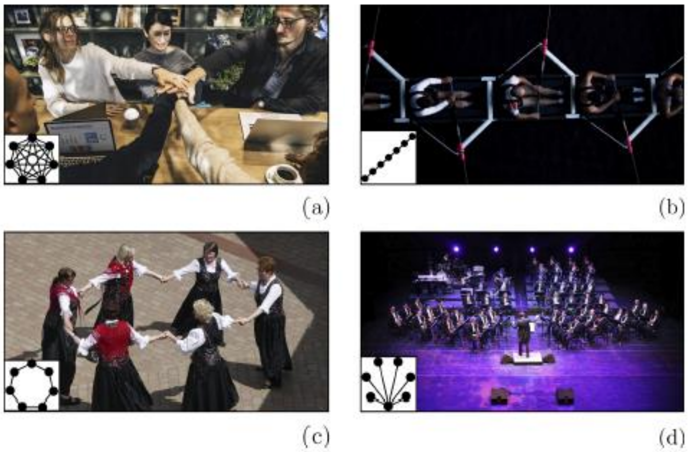


Figure 3: Four topologies during familiar human group cooperation situations, with various coupling modalities. (a) Complete graph: an ordinary organization during everyday working meetings; (b) Path graph: often present in sports, for instance in team rowing where partners are mechanically and visually coupled to two neighbours, except for the first and last rowers; (c) Ring graph: a common structure in many popular dances or among children at play (round dance); (d) Star graph: typical of musical ensembles, for instance when orchestra members are visually coupled only to the director.

Namely, the four topologies were Complete graph, Path graph, Ring graph and Star graph, In Experiment 2, homogeneity among the players was manipulated at a different scale, by comparing groups of novices with groups of certified dancers. For homogeneity, we predicted that similarity would strengthen synchronization, irrespective of graph topology (Experiment 1, see Figure 3), and that dancers would maintain a more solid synchronization regime compared to non-dancers (Experiment 2). For topology, we expected that Complete and Star graphs, that were observed to maximize synchronization metrics during visual contact, would still be associated with higher levels of coordination after visual interruption. Furthermore, we predicted that a stronger memory effect would be present in the case of higher homogeneity between participants (similar pendulum frequencies in Experiment 1 and dancers in Experiment 2) and in graphs producing higher perceptual exchanges (Complete and Star graphs). Experimental results are summarised on Figure 4 and 5.

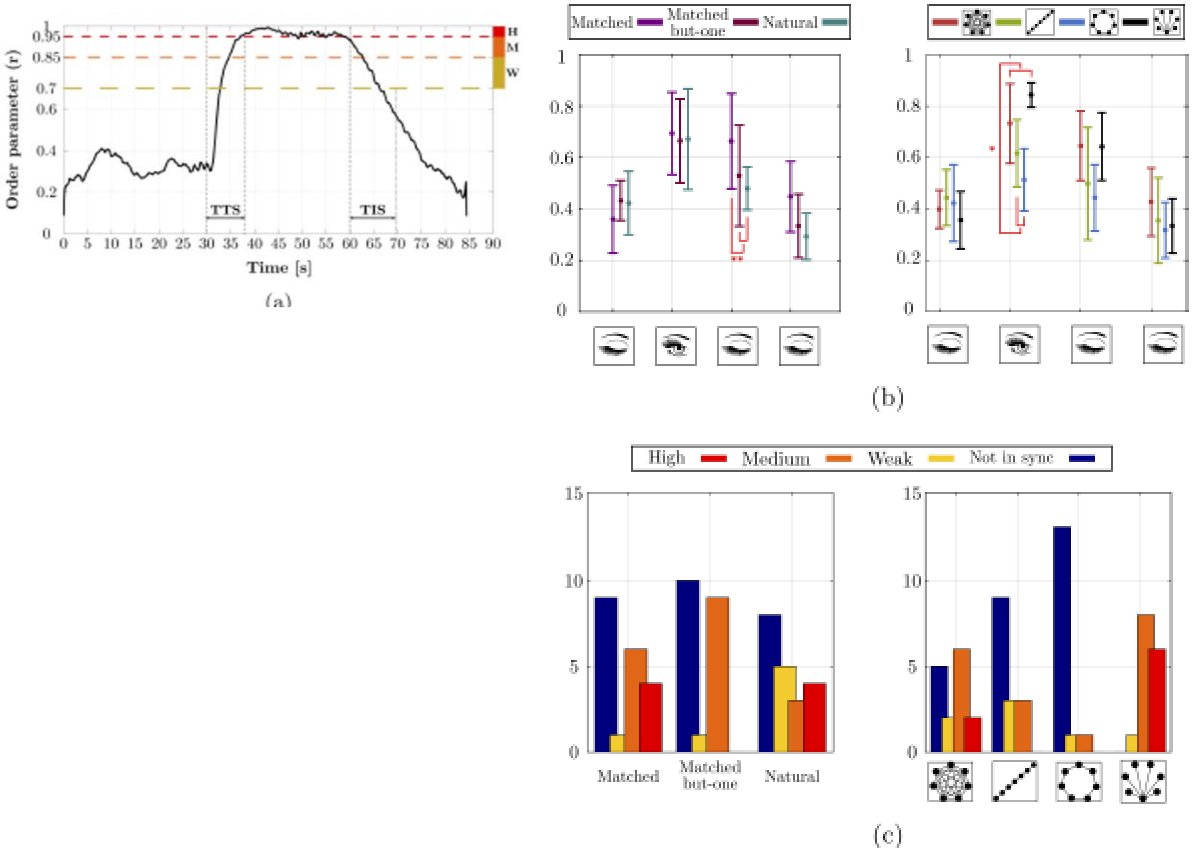


Figure 4: Main results of Experiment (a) a representative example of phase synchronization r across periods of absence and presence of visual coupling (Time To Sync TTS = 7:49 s and Time In Sync TIS = 9:78 s); (b) mean and standard deviation of phase synchronization across homogeneity (left panel) and topology conditions (right panel), n = 240; (c) distribution of phase synchronization levels for Similarity (left panel) and Topology (right panel), n = 60.

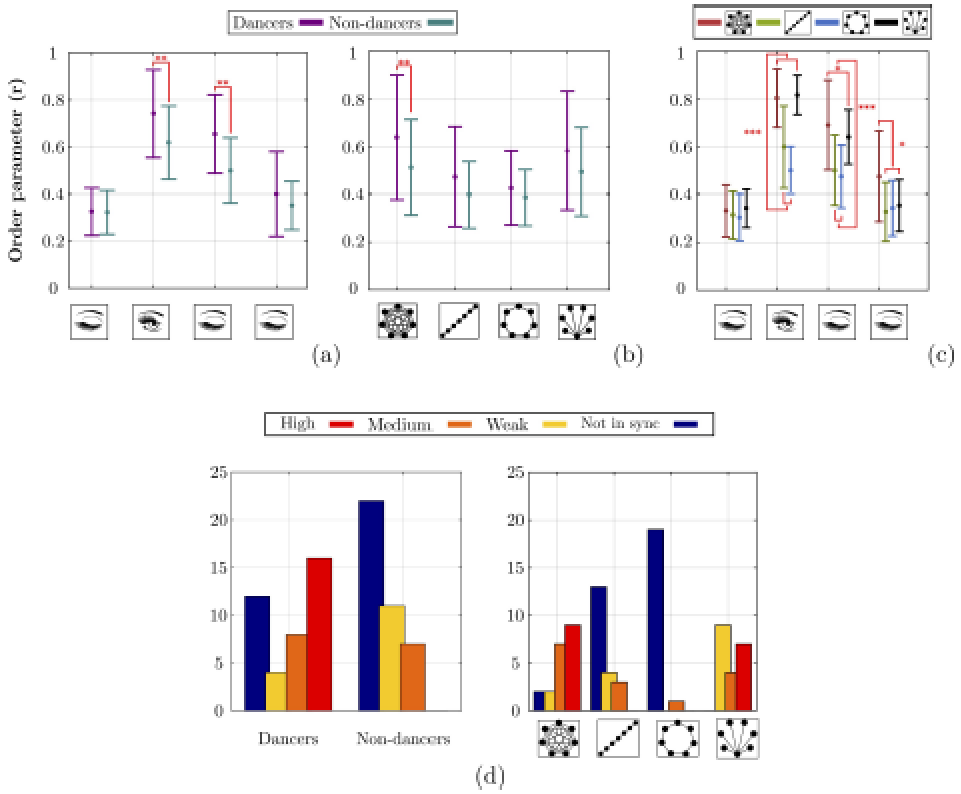


Figure 5: Main results of Experiment 2. Mean and standard deviation of Phase synchronization r in Experiment 2 as a function of (a) VisionExpertise, (b) ExpertiseTopology, (c) VisionTopology), n = 320; (d) distribution of phase synchronization levels across categories of robustness for Expertise (left panel) and for Topologies (right panel), n = 80.

We showed that our ability to move in unison is strongly influenced by our spatial configuration, similarity in behaviour, expertise and amount of visual exchange. In two experiments in which these factors, as well as their key interactions, were manipulated, we demonstrated that Complete and Star graphs were the most solid topologies prone to facilitating synchronized behaviours, reinforced by inertial homogeneity between participants and their expertise in perceptuo-motor synchronization. Importantly, we also demonstrated that group synchronization can be maintained for a certain amount of time (about 7 seconds) after informational exchanges have been interrupted, again more so in the two dominant topologies, and in a stronger way for experts. We investigated the origin of this effect by modelling our behavioural results with a simple ON-OFF dynamical model consisting in switching off the visual coupling and letting the individual dynamics relax to the initial oscillation frequency. This Static Coupling model was sufficient to partially capture our data. However, a memory effect had to be introduced in the model to account for the marked persistence of synchronization in eyes closed for two of the three homogeneity conditions, as well as for the coordination experts. Taken altogether, these results help to better understand why behavioural cohesion is easier to maintain when perceptual exchanges are lost, more so in Path and Ring spatial configurations, and how perceptuo-motor expertise can reinforce this cohesion. How these multiple and co-existing configurations, within and across our senses, modulate our collaborative behaviours, in more naturalistic settings (richer perceptually and socially than pendulum operation in a laboratory), remains largely unknown and constitute a promising avenue for future research we aim to develop during EnTimeMent project.

References

Damm, L., Varoqui, D., De Cock, V. C., Dalla Bella, S. & Bardy, B. Why do we move to the beat? a multi-scale approach, from physical principles to brain dynamics. Neurosci. & Biobehav. Rev. 112, 553–584 (2019).

Kelso, J., Scholz, J. P. & Schöner, G. Nonequilibrium phase transitions in coordinated biological motion: critical fluctuations. Phys. Lett. A 118, 279–284 (1986).
Nordham, C. A., Tognoli, E., Fuchs, A. & Kelso, J. S. How interpersonal coordination affects individual behavior (and
vice versa): Experimental analysis and adaptive hkb model of social memory. Ecol. Psychol. 30, 224–249 (2018).

Oullier, O., De Guzman, G. C., Jantzen, K. J., Lagarde, J. & Scott Kelso, J. Social coordination dynamics: Measuring
human bonding. Soc. Neurosci. 3, 178–192 (2008).

Schmidt, R. C., Carello, C. & Turvey, M. T. Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. J. Exp. Psychol. Hum. Percept. Perform. 16, 227 (1990).

Strogatz, S. Sync: The emerging science of spontaneous order (Penguin Group, New York NY, USA, 2004).

2.3.5 Tracking the leader: gaze behaviour in group interactions

For a full description please see: Capozzi F., Beyan, C., Pierro, A., Koul, A., Murino, V., Livi, S., Bayliss, A. P., Ristic, J., Becchio, C. (2019). Tracking the Leader: Gaze Behavior in Group Interactions. *iScience*;16:242-249. doi:10.1016/j.isci.2019.05.035

It is commonly believed that leadership is reflected in visual behavior. However, little is known about how leadership shapes gaze dynamics during real-world interactions (Capozzi and Ristic, 2018; Koski et al., 2015; Risko et al., 2016). One major reason is the lack of tools to study the gaze behaviours of multiple agents in unconstrained settings.

In this study, we developed a novel tripartite method combining A) computer vision methods for remote gaze-tracking, B) a detailed taxonomy to encode the implicit semantics of multi-party gaze features, and C) advance machine learning methods to establish statistical dependencies between leadership and group visual behaviour during group discussion (Figure 6).

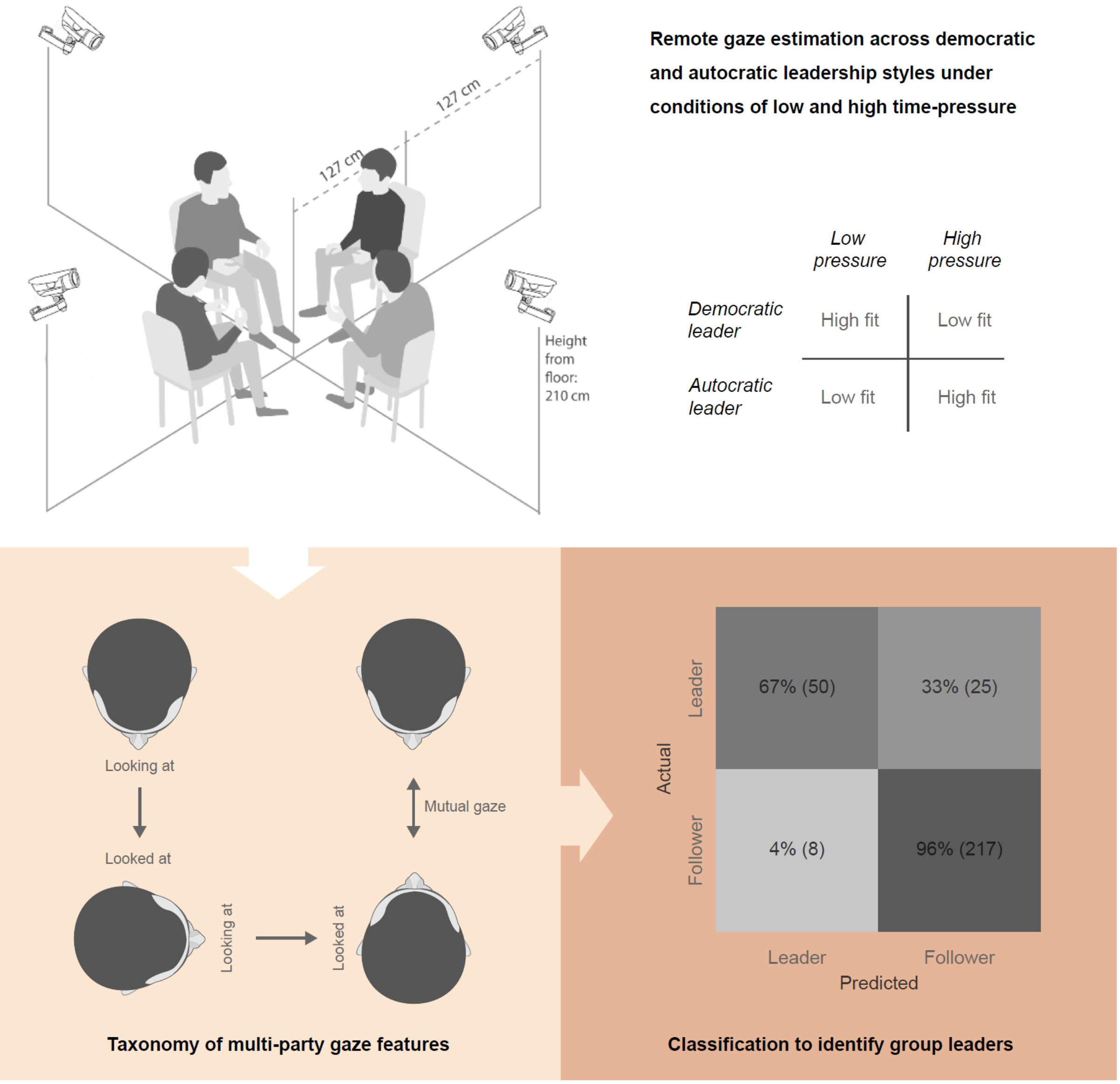


Figure 6: Depiction of gaze estimation methods applied in the study and result matrix.

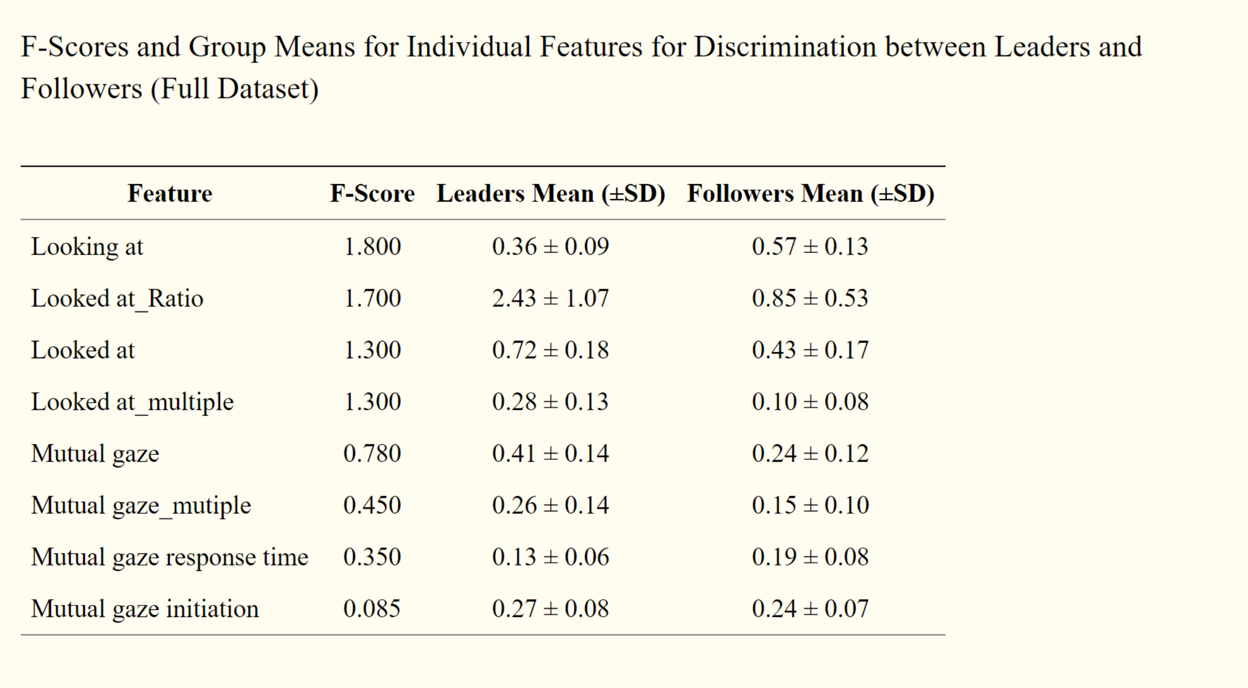
The basic idea of our approach was to conceptualize multi-party gaze features as patterns and to treat the analysis as a pattern classification problem: can the semantics of group visual behaviour reveal the leader among group members? This is the first question we addressed in the study described here. The second question is whether the relationship between gaze behaviour and leadership generalizes across leadership styles and situational conditions – in other words, whether gaze behaviour can serve as a general marker of leadership.

Drawing on ideas from social psychology, we analysed gaze-based interaction dynamics in four leadership settings resulting from the orthogonal manipulation of leadership style (i.e., Democratic vs. Autocratic) and situational condition (i.e., Low time-pressure vs. High time-pressure). Democratic leadership is expected to be more effective under situational conditions of low time-pressure, whereas autocratic leaderships is expected to be more effective under situational conditions of high time-pressure. The orthogonal manipulation of leadership styles and situational conditions resulted in two high-fit conditions (Democratic - Low time-pressure, Autocratic - High time-pressure) and two low-fit conditions (Democratic - High time-pressure, Autocratic - Low time-pressure) (Figure 6). Each group, composed of one designated leader and three followers, was assigned a survival task to solve within a limited time.

First, using a method for automatically estimating the Visual Focus of Attention (VFOA), we determined ‘who looked at whom’. Then, we established a detailed taxonomy of multi-party gaze behaviours and, combining the VFOA of individual group-members, reconstructed the gaze-based interaction dynamics. Next, we probed the actual association between leadership and gaze patterns by asking whether a pattern classification algorithm (SVM) could discriminate between leaders and followers among the group-members (Koul et al., 2018).

Our approach was successful and extremely revealing. We found that social gaze behaviour distinctively identified group leaders. With a cross-validated accuracy of 89%, classification performance was well above the .50 chance level (95% CI = .85, .92; Kappa = .68; Sensitivity = .86; Specificity = .90; F1 = .75; $p < .001$) (Figure 6). To investigate which features were more effective for the classification task, we next computed F-scores (see “Leader classification analysis” in Transparent Methods). F-score provides a measure of how well a single feature at a time can discriminate between different classes. The higher the F-score, the greater the ability of a feature to discriminate between leaders and followers. Table 1 provides an overall view of the discriminative power of each visual feature. Overall, F-scores suggest that leaders looked less at others and, conversely, were looked at more as compared with followers. Also, leaders were involved in and caused more episodes of mutual engagement, relative to followers. The time taken by another group member to respond to the initiation of mutual engagement was also less for leader-initiated episodes compared to follower-initiated episodes.

Table 1



F-Scores and Group Means for Individual Features for Discrimination between Leaders and Followers (Full Dataset)

Feature	F-Score	Leaders Mean (±SD)	Followers Mean (±SD)
Looking at	1.800	0.36 ± 0.09	0.57 ± 0.13
Looked at_Ratio	1.700	2.43 ± 1.07	0.85 ± 0.53
Looked at	1.300	0.72 ± 0.18	0.43 ± 0.17
Looked at_multiple	1.300	0.28 ± 0.13	0.10 ± 0.08
Mutual gaze	0.780	0.41 ± 0.14	0.24 ± 0.12
Mutual gaze_mutiple	0.450	0.26 ± 0.14	0.15 ± 0.10
Mutual gaze response time	0.350	0.13 ± 0.06	0.19 ± 0.08
Mutual gaze initiation	0.085	0.27 ± 0.08	0.24 ± 0.07

With a similar logic, we applied MVCC to test generalization across situational conditions. We trained a linear SVM on gaze patterns recorded under high fit situational conditions (i.e., democratic leaders working in a low time-pressure condition and autocratic leaders working in a high time-pressure condition), and then tested it on group interactions under low fit situational conditions and vice-versa. Cross-classification performance was once again well above the .50 chance level, reaching 94% and 85% for train-high fit and test-low fit (95% CI = .89, .97; Kappa = .83; Sensitivity = .92; Specificity = .94; F1 = .87; p < .001) and train-low fit and test-high fit (95% CI = .78, .91; Kappa = .54; Sensitivity = .82; Specificity = .86; F1 = .63; p < .001), respectively. Collectively, these data show that multi-party visual behaviour supports identification of group leaders across leadership styles (i.e., democratic, autocratic) and situational fit conditions (i.e., high fit, low fit).

To our knowledge, this is the first study that attempts to provide a full characterization of the relationship between leadership and social gaze behaviour during natural group interactions. The novel method utilized in the current study demonstrates that gaze-based group behaviours distinctively identified leaders during natural group interactions. Leaders were looked at more, looked less at others, and elicited more mutual gaze. This pattern was observed over time regardless of leadership style and situational condition, suggesting that gaze can serve as *a general marker of leadership*. Together with previous findings on body movements and paralinguistic behaviours , these results demonstrate the significance of non-verbal cues for leadership identification. We expect that future empirical and modelling studies will investigate whether and how different (and possibly correlated) non-verbal features contribute to leader classification. In addition, we anticipate that these findings will inspire new research questions and real-world applications spanning a variety of domains, from business management to surveillance and politics.

References

Capozzi, F., Ristic, J., 2018. How attention gates social interactions. Ann. N. Y. Acad. Sci. 1426, 179–198.

Koul, A., Becchio, C., Cavallo, A., 2018. PredPsych: A toolbox for predictive machine learning-based approach in experimental psychology research. Behav. Res. Methods 50, 1657–1672.

Koski, J.E., Xie, H., Olson, I.R., 2015. Understanding social hierarchies: The neural and psychological foundations of status perception. Soc. Neurosci. 10, 527–550.

Risko, E.F., Richardson, D.C., Kingstone, A., 2016. Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. Curr. Dir. Psychol. Sci. 25, 70–74.

2.3.6* Capturing human movement and shape information from small groups to extract expressive and social features - using marker-less techniques

*Study is not originally listed in the D1.2

Human beings exhibit phenomenal capabilities in synchronizing joint actions and coordinating at the inter-personal level in a non-verbal manner. This is observed specifically in musical ensembles where co-performers are seen to coordinate their movements effortlessly (Bishop, 2018). Perhaps the most natural response to music is to move and synchronize to the rhythmic elements in and inter-twined in music. When listening to music, we tend to raise our hands, tap our feet, dance, and shake our heads. It has been observed in a musical ensemble, that when a musical piece is being played, there are parts or phrases, basis which the members tend to coordinate their movements with the rhythmic behavior of other group members.

We, as individuals, usually have a general feeling of entrainment, but almost unknowingly, yet spontaneously, we move to music being played around us (Bishop & Goebl 2018a). We may react genuinely to music, and sometimes showcase unique movements in response to what is being heard (Luck et al. 2010; Vuoskoski et al. 2011). The mix of music and the corresponding movement seems to trigger a social bonding effect. These movements or bodily gestures tend to convey certain subtle messages, and this conveyance of messages is crucial for the co-creation of a musical piece, and musicians are seen to perpetually move during a performance to augment the creation of sound, express themselves, communicate with their fellow group members, and transition into states of synchronization (Wanderley, 2001; Bishop & Goebl 2018b). This coordination often takes place in the context of different musical texture that vary in terms of whether there is a clear hierarchy with a leader playing the melody while others play the accompaniment.

Music permits the study of these subtle facets of the human body, using current state-of-the-art methods, while producing minimal noise during experiments. This noise is the random variability one may find in a signal. One way to study such an area is extracting the signals that emerge out of human movements. But when one must perform experiments in-the-wild, noise is a major concern – which adds to the numerical complexities. Another benefit of experimenting with music is that it allows research to be carried out with more care, attention to detail, and control. Music ensemble performances are in turn special examples of joint actions with key advantages. In our experiments, we analyze each of the videos in phrases. These phrases are units of information at relatively long musical timescales. Interpersonal coordination measures using a windowing approach captures shorter timescales. During our analysis we examine different positions of the phrase (start, middle, end). With this, we explore the opportunity to answer questions related to multiple timescales – especially how coordination at shorter timescales changes over the course of longer timescales.

We propose a computational model to compute the synchronization of dyadic pairs in a musical ensemble using marker-less computer vision techniques. Our methods involve the use of human pose estimation algorithms. The human body in pose estimation algorithms is looked at as a system with many elements. Each of these elements, called a key-point, is tracked in a consistent manner through the visual sequences. Eventually, on discovering the position of a group of anatomical joints such as elbows, nose, shoulders, knees, etc., these key-points help identify a blue- print where a skeleton-like structure of the body can be super-imposed. The algorithm then proceeds with completing a pair-wise connection between these key-points to provide a human skeletal structure, super-imposed on the input, known as a Pose. On implementing these algorithms on our dataset, we receive an output as a json file which contains the coordinates of essential body joints.

Interestingly, in a musical ensemble, due to being seated, musicians communicate with each other not using speech, but head and other upper-body movements. They can interact with each other with visual or audio cues. This interaction usually conveys a message to initiate a musical piece at a certain time and could also convey how musical notes need to be played. Thus, we focus on investigating techniques for the automatic analysis of synchronization by tracking the movement of the human head (Yokozuka et al. 2018). The head is tracked by making use of the coordinates of the human nose as provided by the pose estimation algorithm. The kinematic information extracted from the head movements of the performers are then subject to certain signal processing techniques which also find use in neuroscientific research activities to assess connectivity and synchronization of different regions in the brains. We use the data to compute dyadic synchronization between the performers using a metric called the Phase-Locking Value.

The dataset used for these experiments consist of videos from concert performances by the Omega Ensemble, a professional chamber music group from Australia. Each of the videos have been annotated as determined by a musicological analysis based on the published score. For our research the concert video on which experiments were performed was “Brahms Clarinet Quintet” in B minor (Op.115) written in 1891. This work has three movements, which are each designated as a “piece” in the analysis below. Each of these videos being experimented upon have 5 participants.

The goal of our study is to investigate factors that influence the quality of interpersonal coordination between group members: 1. Strength; and 2. Directionality of interpersonal coupling in musical ensemble performance.

Figure 7: Image from a musical piece composed by Johannes Brahms

The specific aim of this analysis is to test how strength and directionality of coupling are influenced by two factors: 1. Position within the musical phrase (Start, Middle, and End); and 2. Musical textures (Homophonic and Polyphonic).

Previous research (Schögler, 1999; Keller et al. 2014) suggests that coupling may be stronger at the beginning and end of phrases than in the middle. The expectation is that when leadership is not assigned there is better interpersonal coordination. This would suggest that coupling strength will be stronger when leadership is distributed – stronger in polyphonic than homophonic texture conditions (Novembre et al. 2015; Noy et al. 2011; Varlet et al. 2020). Also, the presence of a leader may be more influential at the beginning and ending of phrases than in the middle, in which case we would expect a statistical interaction of the two factors. Our study is strongly based on agreements that performers in a musical ensemble produce more coordinated gestures, and by finding the dyadic synchronizations using performer head – movements, we get closer to answering some of these open-ended questions (King et al. 2011). Using non-intrusive techniques, we also present a new computational method to test and present results on the above hypothesis.

The inter-personal coordination is observed by obtaining the dyadic synchronization on selecting a dyad combination of two co-performers in a musical group. For a set of 5 participants, we can have a total of 10 combinations, and a total of 26 video samples have been studied. This includes deciding within the pairs who shall be the melody and accompaniment for between subject effects. This was established on the basis of the musicological analysis, which classified whether each part was playing the melody or accompaniment in the musical score. We then decided that texture will be a between subjects' factor in the Repeated Measures ANOVA because we did not have control over the textures since they were specified in the musical score. Also, no two people are playing the same part – thus also permitting the inclusion of factors such as instruments and understanding their individual effects on a musical phrase.

Our results investigate and demonstrate the variation in body movements during different kinds of musical textures, and delicate transmission of messaging at multiple temporal scales. The Phase Locking Values (PLV) for all pairs were entered into Repeated Measures ANOVA to test for effects of position in the phrase, texture, and pair.

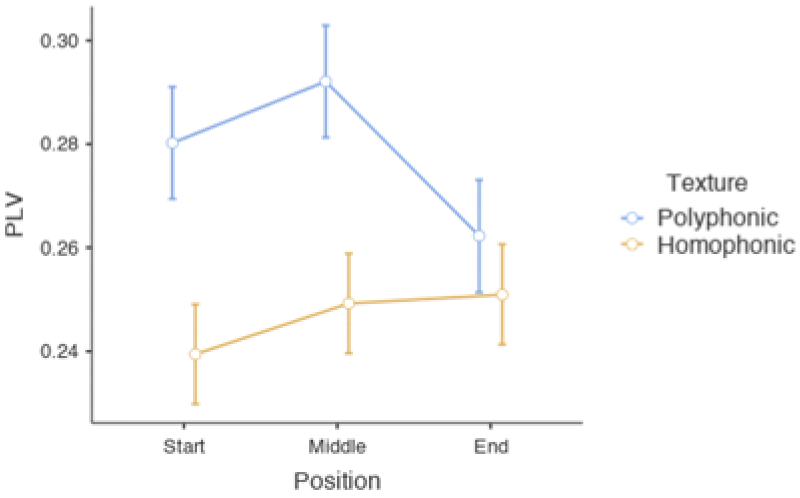


Figure 8: Measure output for Position x Texture

From Figure 8, we observe how the PLV begins at a lower value in both textures, Polyphonic and Homophonic. These PLVs while they start out higher, it tends to rise until the middle of the phrase, and then begins to drop in value towards the end of a musical phrase. It is typically seen from the data plotted that polyphonic textures have stronger coupling than homophonic textures.

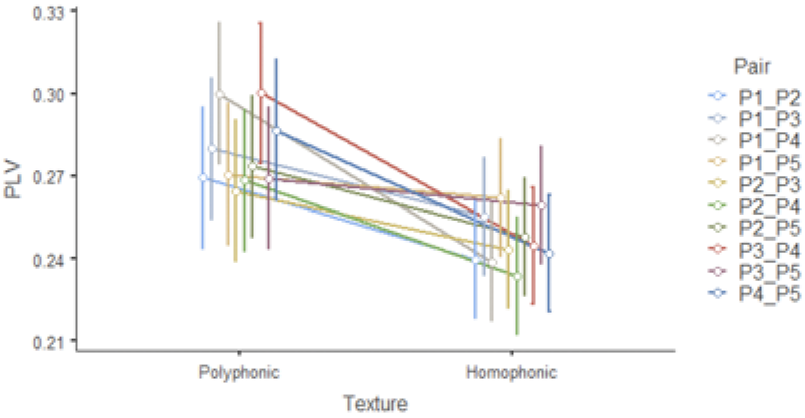
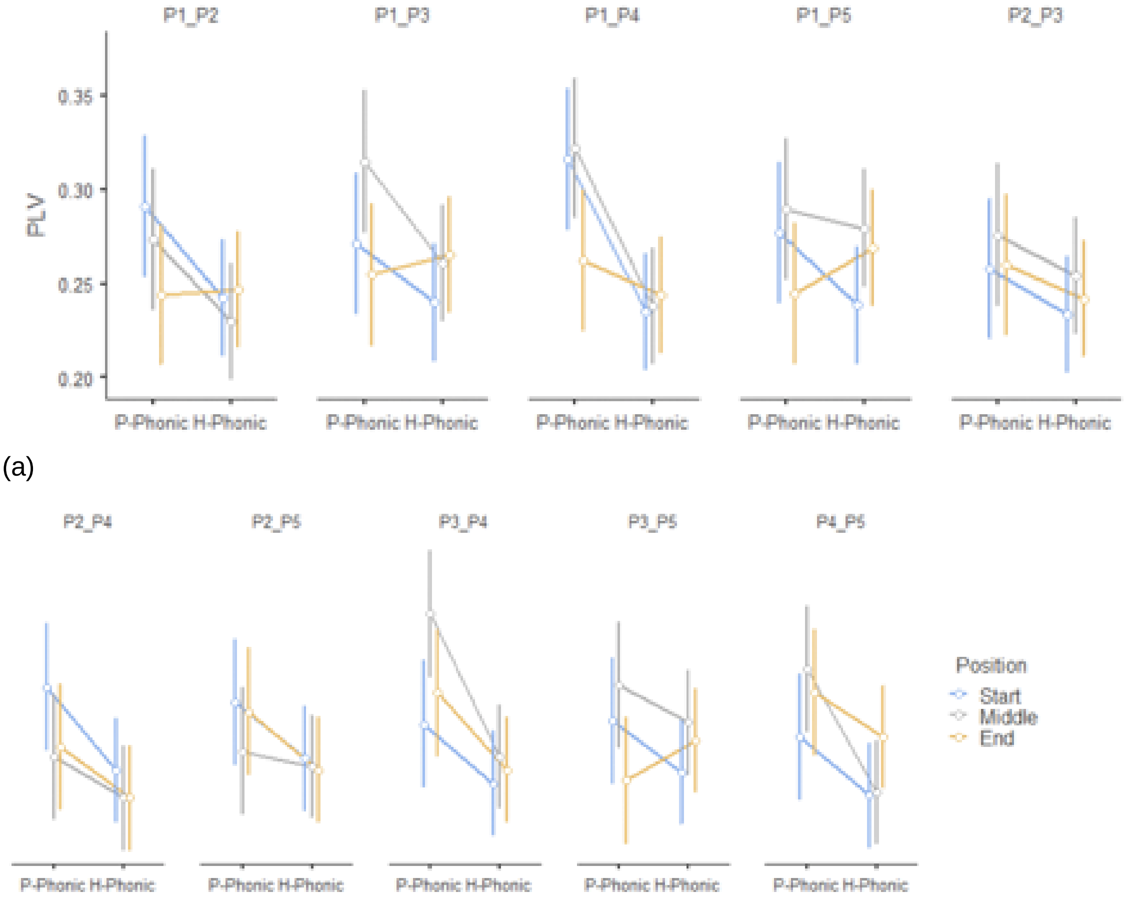


Figure 9: Measures output for Texture x Pair

There were 5 players that synchronized their notes in both the homophonic and polyphonic textures. As seen in Figure 9, pertaining to a pair – wise analysis, we notice that all pairs show a higher level of synchronization in polyphonic textures. We see a sharp or marginal drop in all pairs of performers, suggesting that there may exist sufficient conditions to initiate a leader and follower relationship. This could be because the four performers in a homophonic texture are coupled to the melody player, whereas in a polyphonic texture the coupling is evenly distributed across all performers.



(a)

(b)

Figure 10: Measure output for Texture x Position x Pair

In Figure 10, P-Phonic stands for polyphonic while H-Phonic stands for homophonic, and the results have been arranged in a pairwise order. They depict the results of marginal means of the position as well. We observe that nearly all pairs are seen to have higher PLV values in the middle section of the phrases. In some examples such as P2_5, P2_4, and P1_2, this may differ because while there may be head movements involved, our video is being captured from the front. Thus, the ones seated on the extreme left and extreme right will exhibit more numerical movement due to being seated in a different orientation which can capture better trajectories. Additionally, the drop in Phase Locking Values at the end of the phrase for polyphonic textures could indicate that one of the instruments takes over as leader at that point, thus making it like a homophonic texture and reducing the symmetry of coupling across the ensemble.

We found statistically significant main effects of Position, *F*(2, 400) = 4.657, p=0.01, and Texture, *F*(1, 200) = 34.689, p<0.001. A significant effect of Piece was also observed, *F*(2, 200) = 54.500, p<0.001, but this is currently beyond our theoretical interest. Additionally, the two-way interaction between Position and Texture is statistically significant, *F*(2, 400) = 6.658, p=0.001. This indicates that there is a matching effect between the Position and Texture in musical phrases that have been analyzed – which is in line with the specific aim of our hypothesis and confirms the reliability of the overall pattern of results described in the sections above.

Thus, based on our results, we can say that the effect of the texture changes with respect to the position of the phrase, further helping us study the relationship between musical textures and overall coupling strengths of performers over the course of a musical phrase.

References

Bishop, L. (2018) “Collaborative musical creativity: How ensembles coordinate spontaneity,” Frontiers in psychology, vol. 9, p. 1285.

Bishop, L. & Goebl, W. (2018a) “Beating time: How ensemble musicians’ cueing gestures communicate beat position and tempo,” Psychology of music (46)1, 84– 106.

Bishop, L. & Goebl, W. (2018b) “Communication for coordination: gesture kinematics and conventionality affect synchronization success in piano duos,” Psychological Research, vol. 82, no. 6, 1177–1194.

Luck, G. Saarikallio, S., Burger, B., Thompson, M., and Toiviainen, P. (2010). “Effects of the big five and musical genre on music-induced movement,” Journal of Research in Personality, vol. 44, no. 6, 714–720.

Keller, P., Novembre, G. & Hove, M. (2014) “Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination,” Philosophical Transactions of the Royal Society B: Biological Sciences (369)1658, 20130394.

King, E., & Ginsborg, J. (2011) “Gestures and glances: Interactions in ensemble rehearsal,” New perspectives on music and gesture, 177–2011.

Novembre, G., Varlet, M., Muawiyath, S., Stevens, C. J. & Keller, P. (2015) “The e-music box: an empirical method for exploring the universal capacity for musical production and for social interaction through music,” Royal Society open science (2)11,150286.

Noy, L., Dekel, E., & Alon, U. (2011) “The mirror game as a paradigm for studying the dynamics of two people improvising motion together,” Proceedings of the National Academy of Sciences (108) 52, 20947–20952.

Schögler, B. (1999) “Studying temporal co-ordination in jazz duets,” Musicae scientiae (3)1, 75–91.

Varlet, M., Nozaradan, S., Nijhuis, P., & Keller, P. (2020) “Neural tracking and integration of ‘self’ and ‘other’ in improvised interpersonal coordination,” Neuro Image (206), 116303.

Vuoskoski, J. K., Thompson, W. F., McIlwain, D. and Eerola, T (2011) “Who enjoys listening to sad music and why?,” Music Perception (29)3, 311–317.

Wanderley, M. (2001) “Quantitative analysis of non-obvious performer gestures,” in International Gesture Workshop, 241–253.

Yokozuka, T., Ono, E., Inoue, Y., Ogawa, K.-I. & Miyake, Y. (2018) “The relationship be- tween head motion synchronization and empathy in unidirectional face-to-face communication,” Frontiers in psychology 9, 1622.