D3.6

Data acquisition analytic tools for complex actions

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

Distribution level	[PU][1]
Due date of deliverable	Month 30
Actual submission date	July 2021
Deliverable number	3.6
Deliverable title	Data acquisition analytic tools for complex actions.
Type	
Status & version	
Number of pages	21
WP contributing to the deliverable	3
WP / Task responsible	EuroMov
Other contributors	ALL
Author(s)	EuroMov, IIT_GE, IIT_FE, UNIGE, UCL, Durham
EC Project Officer	Christiane WILZECK
Keywords	Computational models, Software libraries, Movement analysis and prediction, Machine learning; Motion Capture

Abbreviations

EU	European Union
EC	European Commission
PU	Public
WP	Work Package

Table of Contents

1. Introduction

This deliverable reports on the progress on the research conducted between M1-M30 of the EnTimeMent project with regards to the data acquisition tools and analytic (hardware and software) during complex, multi-agent action execution and observation.

This deliverable is strongly connected to deliverable D3.1 (Phase 1) and deliverable D3.5 (Joint Action) which is about the hardware and software tools used for acquiring and analysing data in the context of single and joint actions. As such, this document focuses only on the tools of sub-projects specific for complex action (i.e., overlapping whole body coordination in a multi-agent scenario and group synchronization experiments) and contains references, also in organisation (numbering) of the content, to other deliverables reporting on the D1.7 Models and Algorithms; D2.2; Results on prediction in action execution and observation – Phase 2, and Research Requirements D1.2).

2. Data Acquisition Tools

This section describes the software (processing pipelines), sensors, and other devices used for capturing data in four main settings of the EnTimeMent project: controlled lab settings, unconstrained lab settings, musical performance settings, and everyday (e.g. home) settings.

Please note that the numbering of the subsections refers to numbering used for experimental work in D1.2 Research Requirements, to maintain continuity throughout the project deliverables.

2.1.10 Movement qualities in music performance (Action recognition in Indian classical singers) (UDurham)

Introduction

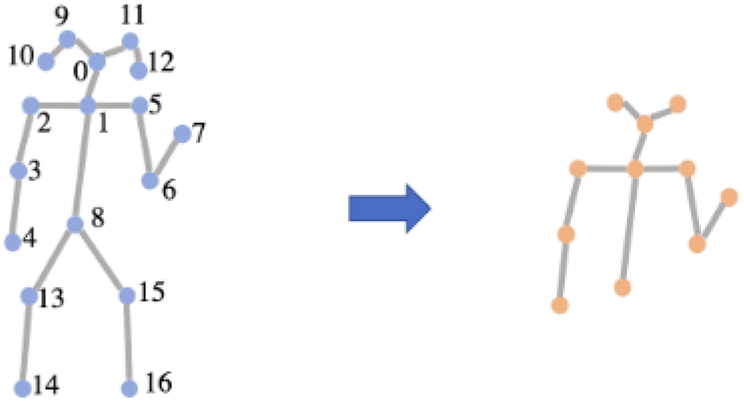
In order to explore the relationship between bodily movement and musical expression, we extract human pose data from videos of Indian classical singers (in the Hindustani khyal genre). We use the resulting movement data to train action recognition models to classify the musical content (the raga) and musicians.

Information is derived from a specially recorded video dataset of solo raga recordings by three professional singers each performing the same nine Hindustani ragas, a smaller duo dataset (one singer with tabla accompaniment) as well as recordings of real concert performances by the same singers. Movement information is extracted using pose estimation algorithms, both 2D (OpenPose, Cao et al. 2021) and 3D (Lifting from the Deep, Tome et al. 2017). A two-pathway convolutional neural network structure is proposed for skeleton action recognition to train a model to classify 12-second clips by singer and raga. The model is capable of distinguishing the three singers on the basis of movement information alone, and within each singers' corpus, of classifying the ragas with up to 49% accuracy.

Method

Pose data is extracted and postprocessed: this involves selection of relevant body parts (ignoring leg data as the musicians are sitting cross-legged on the ground), interpolation of missing data points and smoothing as well as normalization to avoid biasing the model by the video resolution and the framing of the singers. Since the differences in gesturing between ragas, and between musicians, are much less distinguishable than movements such as walking, jumping and sitting, to enhance the signature that may help to identify different classes, movement velocity is computed from the pose data and used as an independent input.

The MS-G3D module (Liu et al., 2020) is introduced to extract features from the skeleton sequences. The original input is a tensor with dimension. For 2D skeleton, =3 including the normalised horizontal and vertical coordinates, and the confidence value returned by the pose estimation algorithms. For the 3-dimensional skeleton, the depth coordinate is included, thus equals to 4. The number of frames for each short video clip denotes, here it is 300 (12-second clips at 25 fps). is the number of key joints of the singer. Considering that singer is sitting in the performance, 11 out of 17 joints are selected.



To calculate the velocity of the movement, we first denote the coordinates of k-th body part in t-th frame as $(x_{k,t}, y_{k,t}, z_{k,t})$. And then the horizontal, vertical and depth velocity values are defined as

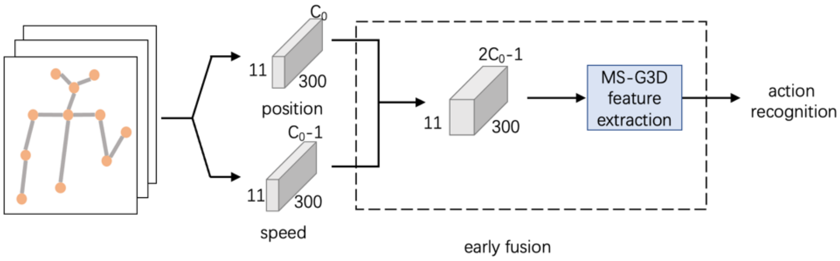
$$\begin{cases} \dot{x}_{k,t} = x_{k,t} - x_{k,t-\Delta} \\ \dot{y}_{k,t} = y_{k,t} - y_{k,t-\Delta} \\ \dot{z}_{k,t} = z_{k,t} - z_{k,t-\Delta} \end{cases}$$

where Δ is the frame interval for calculating the velocity. For the case that $t \leq \Delta$, we simply let

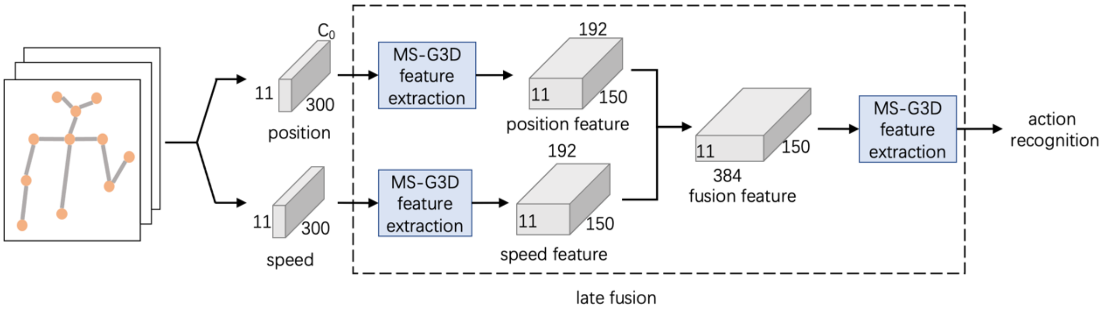
$$\begin{cases} \dot{x}_{k,t} = \dot{x}_{k,\Delta+1} \\ \dot{y}_{k,t} = \dot{y}_{k,\Delta+1} \\ \dot{z}_{k,t} = \dot{z}_{k,\Delta+1} \end{cases}$$

Therefore, the dimension of velocity tensor is $(C_0 - 1, T, K)$.

In order to combine the position and speed information, two fusion strategies are evaluated, termed ‘early fusion’ and ‘late fusion’, and compared to the original MS-G3D model. For the early fusion, the tensors of position and speed are concatenated to form a new tensor with dimension $(2C_0 - 1, T, K)$. This tensor is fed into the MS-G3D model to identify the ragas or musicians.



For the late fusion, a MS-G3D module is introduced to extract the skeleton-based feature, of which the dimension is $(C_1, T/2, K)$. Then features from two channels are concatenated to a new tensor with dimension $(2C_1, T/2, K)$. After that, the fusion feature is fed to another MS-G3D module to predict the score of all the actions that are required to be identified. The illustration of the late fusion model is shown below.



Results of the raga and musician recognition are shown below.

Table 1. Accuracy of raga classification for different musicians (AG, CC, SCh)

	AG	CC	SCh	mean
2D MS-G3D	29.0%	**33.7**%	**49.0**%	37.2%
2D early fusion	35.3%	33.0%	46.4%	**38.2**%
2D late fusion	**37.7**%	25.1%	45.1%	34.6%
3D MS-G3D	28.0%	20.7%	**39.1**%	29.2%
3D early fusion	**30.2**%	**25.4**%	33.3%	**29.6**%
3D late fusion	27.1%	19.5%	34.2%	26.9%

The mean accuracy for the raga classification is about 26.9% ~ 38.2% according to different models, all of which are better than the random guess (1 out of 9 ragas, i.e. c. 11%). The average classification accuracy using 2D pose data is about 8% better than that using 3D pose. The early fusion strategy in the two-pathway method achieves the highest mean classification accuracy.

Table 2. Accuracy of singer classification on the different test set.

	Random solo	Duo/ concert
2D MS-G3D	100.0%	60.5%
2D early fusion	99.8%	**68.8**%
2D late fusion	100.0%	56.9%
3D MS-G3D	100.0%	71.8%
3D early fusion	100.0%	**73.0**%
3D late fusion	100.0%	58.9%

When we randomly divide solo videos into training and test sets and then split the video into short clips, all methods success to identify the musicians. When the model is trained in solo videos, it performs well on identifying the same musician in totally different scenes. The early fusion approach again achieves the best results, but in this case the 3D data gives more accurate classification.

References:

Cao, Z., G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43 (1), pp. 172-186, 1 Jan. 2021, https://doi.org/10.1109/TPAMI.2019.2929257.

Liu, Z., Zhang, H., Chen, Z., Wang, Z., & Ouyang, W. (2020). Disentangling and unifying graph convolutions for skeleton-based action recognition. In Proceedings of *the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 143-152.

Tome, D., Russell, C. & Agapito, L. (2017). Lifting from the Deep: Convolutional 3D Pose Estimation from a Single Image. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR.2017.603

2.1.17 Intersecting action and perception in autism spectrum disorders at the single-trial level (IIT-GE)

Extraction of behavioural indexes of action prediction. We developed a new analytical framework to study intention encoding (how intention information is encoded in movement kinematics) and intention readout (how intention information is read out from visual kinematics) with single-trial resolution. We recently applied this framework to investigate how intention encoding and readout intersect at the single-trial level in autism spectrum disorders. We report on the study in D2.2 (Study 2.1.17 Intersecting action and perception in autism spectrum disorders at the single-trial level – as specified in D1.2).

To determine the dependency of intention (encoding model) and intention choice (readout model) on kinematics over time, we used a logistic regression to estimate the single-trial cumulative probability y(t) (that is, the cumulated evidence) in favour of one intention (e.g., 'to place') as function of the time-dependent kinematic vector in that trial until time t. Specifically, we modelled y(t) as a sigmoid transformation of the sum of two terms: a linear transformation of the kinematic vector K (t), which describes the evidence provided by the single-trial kinematic vector at the current time epoch (t), and a drift term, which describes contribution of the cumulated evidence y(t-1) provided by the kinematic vectors up to the previous time epoch (t-1).

2.2.7 Slow and fast sync (dynamical model and cultural comparison approach) (EuroMov, UDurham)

Development and learning in interaction with the environment, including repeated exposure and interaction with patterns determined by culture, constitute an example of very slow changes, on an individual's lifespan scale, that influence rhythmic skills (Jacoby & McDermott, 2017). Along this line of thinking, we aim at analysing how culture pervades across general rhythm skills and specifically determine elementary synchronization. Our first entry point was the comparison of Indians and Frenchs participants. Data collected this spring, including 15 French and 15 Indian participants, show interesting differences in the way to synchronize to a simple beat. The data collected points at analysing further in follow ups two-time scales of adaptation: Frequency and phase. For definitions and analysis, the approach uses the theoretical framework of coordination dynamics. The basic model is a non-linear model of a self-sustained oscillator (l.h.s.), forced by a periodic function and random noise (r.h.s.):

$$\ddot{x} + \dot{x}^3 - \dot{x} + \dot{x}.x^2 + \omega 0 x = \varepsilon.\sin(\omega.t) + \sqrt{Q}.\xi t \qquad \textit{Eq. 1}$$

It is well known that this model of synchronization obeys the so- called theory of Arnold's tongues (Kelso & DeGuzman, 1988), enabling identifying a priori the determiners of synchronization. From this equation relative phase dynamics can be obtained, bistable dynamics of two stable attractors, synchronization and syncopation, resp. in phase and antiphase (Kelso et al., 1990; Eq. 2):

$$\dot{\phi} = \Delta\omega + asin\phi - \mathrm{bsin}2\phi \ + \sqrt{Q}.\xi t \qquad \textit{Eq.2}$$

Here we study exclusively synchronization, therefore the bistable equation Eq. 2 can be linearized to obtain further meaningful observables.

We ran an experiment examining the hypothesis that the behavioural difference observed between the Indians and French synchronization comes from sensorimotor adjustments evolving at two-time scales, corresponding in short to period or phase adjustments. We aim at i) making this assumption more explicit based on available modelling, and ii) testing explicit predictions from the theory, iii) isolate essential aspects of cultural factors that determine those differences.

Participants

Indians and French participants (N = 15 in each group, 11 men and 4 women, age 22 to 45), all students at the university, right-handed, recruited in Montpellier, were matched in pairs to control for education, age, and musical, or dance, or sports experience. Indians recruited had left India less than 2 years before the experiment, their mother tongue was Indian, their second language English, and they were not fluent in French. Participants gave informed consent before the experiment.

Task

The task was to synchronize as best as possible a tap on the table of the index finger with a sound. 3 trials were completed. The frequency of the sound beats (40ms; carrier frequency 440Hz) sequence was increased every 15 stimuli by 0.3 Hz. The range of the pacing frequency went from 1 to 6.1 Hz.

Data collection

A goniometer was used to collect the index finger position (metacarpophalangeal angle), connected to an A to D card, also used to collect stimuli. To get a good accuracy for determining the temporal center of each auditory beat we collected all signals at 5KHz. A second PC and the sound D to A card was used to display the stimuli.

Data pre-processing

Angular positions were down sampled to 500Hz and low pass filtered at 30Hz with dual pass to negate the phase shift. Stimuli were processed to identify the time of each centre of beat, using a low pass (dual pass) filter and local maxima estimation.

Data processing to measure synchronization

The relative phase between position and beats was estimated taking the value of the Hilbert transform phase of the position at each stimuli onset. Transients (beginning of each plateau) were excluded when calculating mean and dispersion of relative phase. The angular mean and dispersion were estimated and are well defined in stationary behaviours.

Results

The maximal rates at which French and Indian participants were able to synchronize were comparable. However clearly the classical negative mean asynchrony is not observed in the Indians participants (Figure 1).

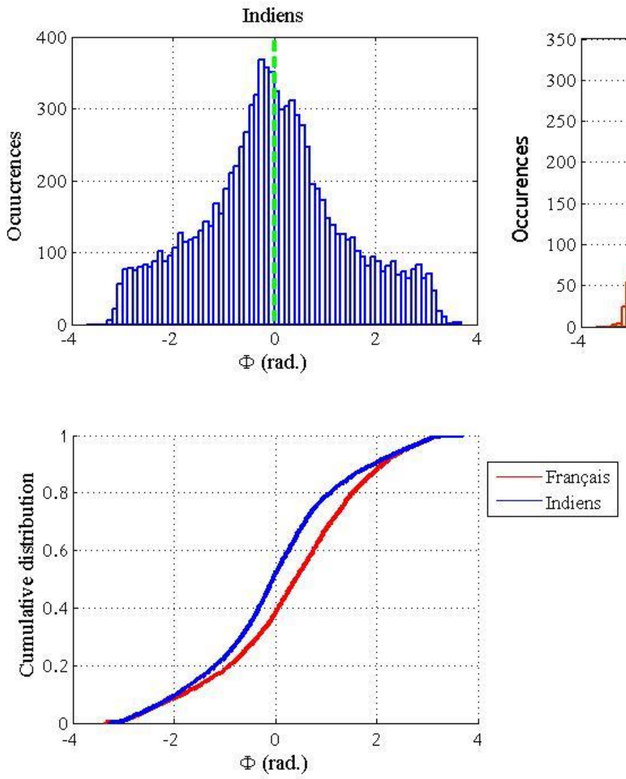


Figure 1. The first task, frequency ranging from 1Hz to 6.1Hz. Histograms of relative phases for all the plateaus for French and Indian participants (N = 9720 values; bin size 0.1 radians). The lower

panel shows the cumulative distributions; a Kolmogorov-Smirnov test on the maximal difference between cumulative distributions confirms a significant difference between the distributions of the two groups. The distribution of French participants is centred toward positive relative phase, while for the Indians participants the distribution is centred on negative values. Please remember that the sign is reversed relative to usual conventions: Positive correspond to a movement advance in time with respect to the stimuli, which is the classic mean negative asynchrony.

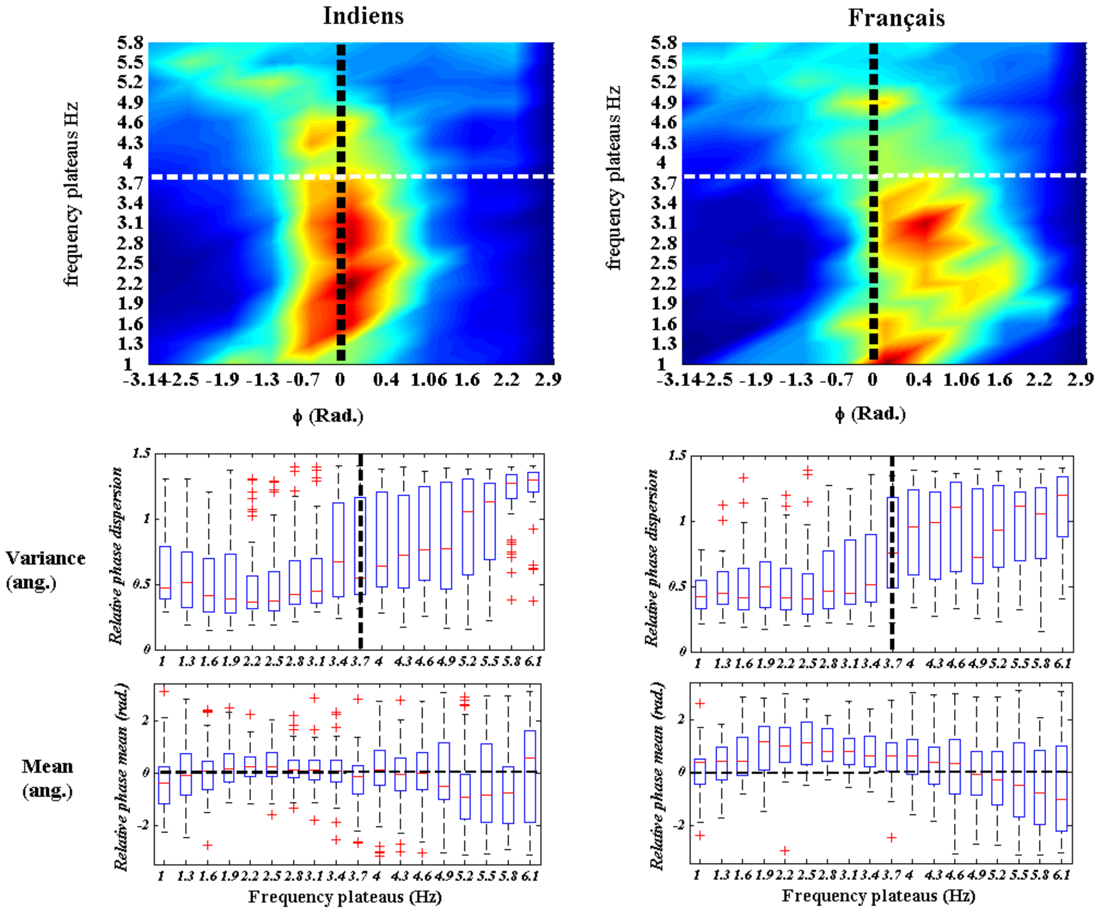


Figure 2.
The first task, frequency ranging from 1Hz to 6.1Hz. On the top row the color coded histograms of the relative phase as a function of the frequency of the stimuli (Red is high occurrences, blue is rare occurrences). On the middle row, the box plot of the angular dispersion (variance) of the relative phase. On the bottom row, the box plot of the mean dispersion (variance) of the relative phase. Please remember that the sign is reversed relative to usual conventions: Positive correspond to a movement advance in time with respect to the stimuli, which is the classic mean negative asynchrony.

References:

Bose, A., Byrne, A., & Rinzel, J. (2019). A neuromechanistic model for rhythmic beat generation. PLoS computational biology, 15(5), e1006450.

Jacoby N, McDermott JH (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. Current Biology, 27, 359-370.

Kelso JAS, DeGuzman GC (1988) Order in time: how cooperation between the hands informs the design of the brain. In: Haken H (ed) Neural and Synergetic Computers. Springer, Berlin Heidelberg New York, pp 180-196.

Kelso JAS, DelColle J, Schöner G. (1990). Action-perception as a pattern formation process. In: Jeannerod M (Ed.). Attention and Performance XIII. Hillsdale, NJ: Erlbaum. pp. 139–169.

Lagarde J, Kelso JAS (2006) Binding of movement, sound and touch: multimodal coordination dynamics. Exp Brain Res. 173, 673-88.

2.3.1. Orchestra violin sections and conductor (IIT-FE; UNIGE)

Introduction

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. To achieve fast inter-individual coordination, individuals may build internal predictive models of other's behavior. In function of the context, the most appropriate motor model is compared with the current observed movement, to generate a prediction error (Friston, Mattout, and Kilner 2011) and update own motor planning (Sebanz, Bekkering, and Knoblich 2006). 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, D'Ausilio, et al. 2016; D'Ausilio, Novembre, et al. 2015).

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 temporal coordination (modes of communication), through different channels of communication (instrumental and ancillary movements) during different playing situations (normal and perturbed).

Participants

A chamber orchestra consisting of 8 violinists (2 sections of four violinists: S1 and S2) and 10 instrumentalists participated in the study along with two professional conductors (C1 and C2). Data were collected from the two violinists' sections and conductors. Each violinists section counted four players.

Apparatus and set-up

Movement data were collected (1000Hz) by using a Qualisys motion capture system equipped with 7 cameras, integrated with the EyesWeb XMI platform: http://www.infomus.org/eyesweb_ita.php (Volpe, Alborno, et al. 2016), including audio and physiological signals (not used here). Each violinist was equipped with a cap on which were placed three passive markers of the Qualisys motion capture system. The positions, based upon the 10-20 electroencephalographic system, corresponded to Pz, F3 and F4. Before starting recording sessions, we ensured that the cap was not moving with musicians' movements and facial expressions. An average of these three markers was taken for further analysis on head movement, to minimize loss of data and interpolation. An additional marker was placed on the bows of the players and on the baton of the conductors. After data tracking by using the Qualysis Track Manager software, the data was exported and analyzed in MATLAB.

Data pre-processing and analysis

We first used the spline method to handle the missing data in the 3D trajectories. The spline method interpolates missing data with continuous third order derivatives. We then computed the magnitude of the acceleration from each 3D trajectory (as done in (D'Ausilio et al. 2012)). Acceleration was chosen because it should be more informative than trajectory and velocity, especially for what regards expressive information transfer. This claim is backed by studies on visuo-motor coordination suggesting that a marked deceleration towards the endpoint of a moving object's trajectory provides more saliency to the timing of this endpoint, and facilitates synchronization with that object (Varlet et al., 2014; Zelic et al., 2016). Each musician time-series on each trial was normalized (to z-scores) and outliers (>6std) were set as absent values (NaN) and interpolated when the gap was smaller than 200 frames (i.e. 2sec). The total percentage of interpolated data was: 4.7% (±1.6).

In the following, we made an empirical dissociation between two modes of communications. We considered as intra-group temporal coordination, the relation between musicians playing the same score. Due to common score, these musicians are engaged in a joint behavior requiring an important degree of temporal coordination. In parallel, we named inter-group complementary coordination, the relation between musicians having different scores, and thus being engaged in a joint behavior requiring an important degree of movement complementarity.

Intra-section temporal coordination: Principal Component Analysis. To evaluate the level of temporal coordination between violinists' movements of each section of violinists (playing the same score), we used a principal component analysis (PCA). PCA is a standard statistical technique generally used to extract a low-dimensional structure from a high-dimensional dataset. Dimensionality reduction method are classically used in the motor synergies field to extract invariant/similar features across time between muscle or kinematic parameters. In particular, PCA has been used to characterize the degree of covariance across time of different body segments in whole-body movements (e.g., reaching (Berret et al. 2009)). Here, PCA was performed on the acceleration profiles of the four violinists of each section (Figure 1, lower panel), windowed and pre-processed in the same way as Granger Causality analysis. Mathematically, the method involves the eigenvalue decomposition of a dataset covariance matrix in order to find the principal directions in the high-dimensional space. For each of the windows, we considered an input matrix composed of 300 rows (temporal frames) and 4 columns (the acceleration profiles of the four violinists in each section) to which we applied the Matlab *princomp* function, after a zscore normalization of the input matrix. The PCA gives four principal components (PC) each written as a linear combination of the initial waveforms (the four violinists' acceleration profile). The variance accounted for (VAF) by the first principal component (noted PC1%) is defined as the ratio between the first eigenvalue and the sum of all the eigenvalues. The VAF represents the degree to which the linear combination associated to each PC is able to approximate the initial dataset. A high PC1% value means that the trajectory in the space of angles is close to a straight line (i.e., all angles were linearly correlated together) while, a low PC1% value indicates that one principal component is not sufficient to describe precisely the trajectories.

Conductor behavior predictability: auto-regressive model's fitting. Following sensorimotor communication literature (Pezzulo et al., 2018 for a review), we evaluated conductor behavior predictability to verify whether intrinsic variability of conductor's behavior was altered by our experimental manipulation. In cooperative joint action tasks, leaders tend to make their movements more consistent over time to help their partner build a predictive model of other's action. An increase in the predictability of (Partner) A translates into a smaller uncertainty when (Partner) B needs to predict future signals coming from A to plan the most appropriate action. We evaluated the level of predictability of conductors' behavior as goodness of fit of the linear autoregressive model computed on the conductor acceleration profile extracted from bow and head data separately. We modelled the conductor acceleration profile via a linear autoregressive model in the same way we computed it for Granger Causality analysis and on the same sliding windows parameters. The optimal order of the model was determined via the Akaike's information criterion and the goodness-of-fit (ARfit) was measured as the sum of squares of the residuals, for each sliding window.

Inter-group complementary coordination: Granger causality analysis. Granger causality analysis was then carried out on the preprocessed acceleration waveforms. According to Granger formalism, a signal X Granger-causes (or G-causes) a signal Y if the past values of X contains information that helps predict Y above and beyond the information contained in the past values of Y alone. Thus, a Granger-causality score (gca) was defined between each pair of musicians as the log-likelihood ratio of the degree to which the prior time series of a musician X (causing variable) contributes to predict the current status of a musician Y (dependent variable), over and above the degree to which it is predicted by its own prior time series while conditional on the remaining musicians time-series (conditional variables). The use of conditional allow to take into account the influence of musicians out of the tested pair to avoid misinterpretation due to multiple sources of information (D'Ausilio et al. 2012). Gca was evaluated (pairwise), every 500 milliseconds on 3-s sliding windows using the "Granger Causality connectivity analysis" Matlab toolbox (Seth 2010). Windows containing more than one third (i.e., 166ms) of absent values were not used in the analysis (less than 5% of the total windows number). The Granger Causality computation is similar to the one used in (Badino et al. 2014; D'Ausilio et al. 2012). From this point, we will represent gca of X on Y by the notations $G_{X\text{->}Y}$ or X->Y.

We were interested in the causality relations between the conductor and each section of violinists (S1 and S2). We performed three different types of Conditional Granger causality computations: (1) Causality between each conductor and violinists of S1 (taken separately): defining as causing variable the conductor, as dependent variable each S1 violinist separately and the other way around [conditional variable: musicians in S2 - taken separately]. (2) Causality between each conductor and violinists of S2 (taken separately): defining as causing variable the conductor, as dependent variable each S2 violinist separately and the other way around [conditional variable: musicians in S1 - taken separately]. (3) Causality between the violinists of S1 and S2 (taken separately): defining as causing variable each S1 violinist separately, as dependent variable S2 violinists separately and the other way around [conditional variable: the conductor]. In these three analyses, we computed gca between each pair of musicians on each 3s window. When the causality between the two variables was significant, we kept the gca value otherwise this value was set to 0. Finally, gca values were averaged across conditional variables, conductors and musicians of same section, to get one value per group (i.e. C->S1, S1->C, C->S2, S2->C, S1->S2, S2->S1). Thus, for each experimental condition, the output matrix consisted of 6 columns (the number of causal relation) and thousands of lines (the number of considered windows).

More details on Granger computation are reported in D3.1.

Statistical analyses

Inter-group and intra-group data did not follow a normal distribution according to normality tests (Kolmogorov–Smirnov) and the variances were also not homogeneous according to statistical tests (Levene). We, therefore, used a two-tail independent samples Welch's t-test (already used on same type of data in (Badino et al. 2014)). In the Welch's t-test the assumption of normality is not critical for large samples (Geary 1947) as it is the case for our data set. More importantly, Welch developed an approximation method for comparing the means of two independent populations when their variances are not necessarily equal (Welch 1947). Because Welch's modified t-test is not derived under the assumption of equal variances, it allows the comparison of two populations without first having to test for equality of variance.

Based on the data extracted in “intra-section temporal coordination”, we made four comparisons for each kinematic parameter: %PC1$_{S1\ NORM}$ vs %PC1$_{S1\ PERT}$, %PC1$_{S2\ NORM}$ vs %PC1$_{S2\ PERT}$, %PC1$_{S1\ NORM}$ vs %PC1$_{S2\ NORM}$, %PC1$_{S1\ PERT}$ vs %PC1$_{S2\ PERT}$. For “conductor behavior predictability” we compared for each kinematic parameter: ARfit $_{NORM}$ vs ARfit $_{PERT}$.

Based on the data extracted in the “inter group complementary coordination”, we made three different set of comparisons, repeated twice (once for head data, once for bow data). (1) For the normal condition, we ran 5 comparisons: C->S1 vs S1->C, C->S2 vs S2->C, S1->S2 vs S2->S1, C->S1 vs C->S2, S1->C vs S2->C. The other possible comparisons were not performed because they were not informative for the study (e.g. C->S1 vs S2->C) or comparing elements of different nature (e.g. C->S1 vs S2->S1). (2) For the perturbed condition, we ran the same 5 comparisons as in (1). (3) Across the two experimental conditions, we ran 6 comparisons: C->S1$_{NORM}$ vs C->S1$_{PERT}$, C->S2$_{NORM}$ vs C->S2$_{PERT}$, S1->C$_{NORM}$ vs S1->C$_{PERT}$, S2->C$_{NORM}$ vs S2->C$_{PERT}$, S1->S2$_{NORM}$ vs S1->S2$_{PERT,}$ S2->S1$_{NORM}$ vs S2->S1$_{PERT}$.

In all these analyses, the p-level was corrected for multiple comparisons with the Benjamini and Hochberg false discovery rate procedure. We reported in the results part the corrected p-value, and the value of the test statistic. We considered as marginally significant the statistical comparison for which the p-value before correction was inferior to 0.05. All analyses were conducted using the Matlab Statistics toolbox (Mathworks Inc.).

References:

Badino, Leonardo, Alessandro D’Ausilio, Donald Glowinski, Antonio Camurri, and Luciano Fadiga. 2014. “Sensorimotor Communication in Professional Quartets.” *Neuropsychologia* 55 (1). Elsevier: 98–104. doi:10.1016/j.neuropsychologia.2013.11.012.

Berret, Bastien, François Bonnetblanc, Charalambos Papaxanthis, and Thierry Pozzo. 2009. “Modular Control of Pointing beyond Arm ’ s Length.” *The Journal of Neuroscience* 29 (1): 191–205. doi:10.1523/JNEUROSCI.3426-08.2009.

D’Ausilio, Alessandro, Leonardo Badino, Yi Li, Sera Tokay, Laila Craighero, Rosario Canto, Yiannis Aloimonos, and Luciano Fadiga. 2012. “Leadership in Orchestra Emerges from the Causal Relationships of Movement Kinematics.” *PLoS ONE* 7 (5). doi:10.1371/journal.pone.0035757.

D’Ausilio, Alessandro, Giacomo Novembre, Luciano Fadiga, and Peter E Keller. 2015. “What Can Music Tell Us about Social Interaction ?” *Trends in Cognitive Sciences* 19 (3). Elsevier Ltd: 1–4. doi:10.1016/j.tics.2015.01.005.

Friston, Karl J, Jérémie Mattout, and James M Kilner. 2011. “Action Understanding and Active Inference.” *Biological Cybernetics* 104 (1–2): 137–60. doi:10.1007/s00422-011-0424-z.

Geary, R C. 1947. “Testing for Normality.” *Biometrika* 34 (3/4): 209–42.

Sebanz, Natalie, Harold Bekkering, and G??nther Knoblich. 2006. “Joint Action: Bodies and Minds Moving Together.” *Trends in Cognitive Sciences* 10 (2): 70–76. doi:10.1016/j.tics.2005.12.009.

Sebanz, Natalie, and Guenther Knoblich. 2009. “Prediction in Joint Action: What, When, and Where.” *Topics in Cognitive Science* 1 (2): 353–67. doi:10.1111/j.1756-8765.2009.01024.x.

Seth, Anil K. 2010. “A MATLAB Toolbox for Granger Causal Connectivity Analysis.” *Journal of Neuroscience Methods* 186 (2): 262–73. doi:10.1016/j.jneumeth.2009.11.020.

Volpe, Gualtiero, Alessandro D’Ausilio, Leonardo Badino, Antonio Camurri, and Luciano Fadiga. 2016. “Measuring Social Interaction in Music Ensembles.” *Philosophical Transactions B* 371 (April): 20150377. doi:10.1098/rstb.2015.0377.

Welch, B L. 1947. “The Generalization of ‘Student’s’ Problem When Several Different Population Variances Are Involved.” *Biometrika* 34 (1–2): 28–35.

Wolpert, Daniel M, Kenji Doya, and Mitsuo Kawato. 2003. “A Unifying Computational Framework for Motor Control and Social Interaction.” *Philosophical Transactions of the Royal Society of London* 358 (February): 593–602. doi:10.1098/rstb.2002.1238.

2.3.4 Similarity of motor signatures across multiple timescales in musical performers: Computing Dyadic Synchronisation using marker-less techniques. (UNIGE)

As illustrated and explained in document D1.7, we are looking at multiple ways of investigating dyadic synchronisation between musicians in an ensemble.

A recent update in our project is to also look at Hilbert-Huang Transform (HHT) as a way to decompose the motion signals into their instantaneous frequency scale. This is being done in addition to Fourier Transform (FFT) to observe differences, if any, in our results.

On using RMPE (Regional Multi-Person Pose Estimation) or AlphaPose (Fang, Hao-Shu, et al. 2017), we are able to extract robust 2D key points that is extracted, and later exploited to study human movement behavior. This data is available in the form of a json file at the end of applying the algorithm. We adopt a 7-step methodology to help obtain phase-locking values using. The diagram below has been modified to include HHT also as a part of the process (Figure 1).

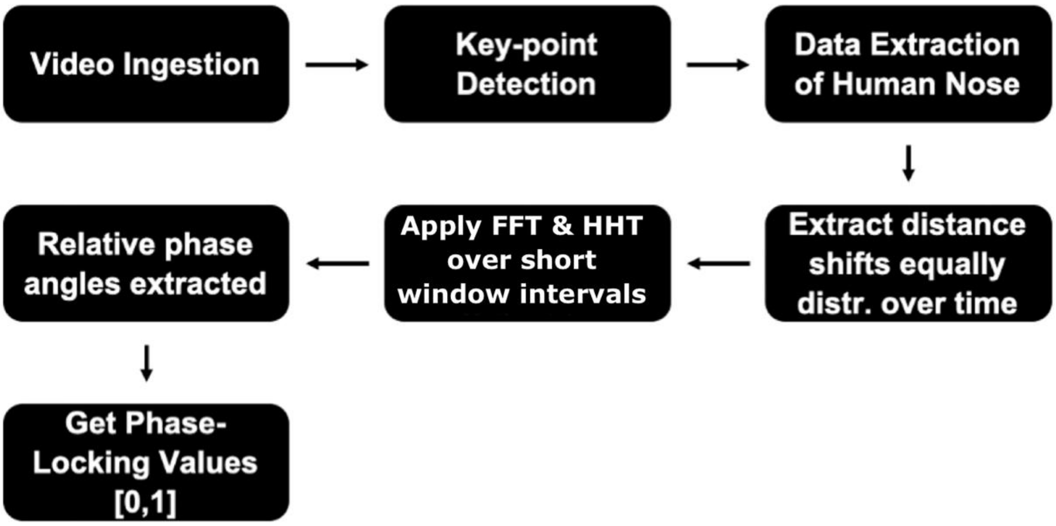

Video Ingestion
Key-point Detection
Data Extraction of Human Nose
Extract distance shifts equally distr. over time
Apply FFT & HHT over short window intervals
Relative phase angles extracted
Get Phase-Locking Values [0,1]

Figure 1.
The Methodology followed for obtaining Phase-Locking Values

On computing the phase-locking values, we then compute the Dyadic Synchronization that emerges during the performances. To further, reference, below is an eight-step pipeline that is utilized to compute the Dyadic Synchronization.

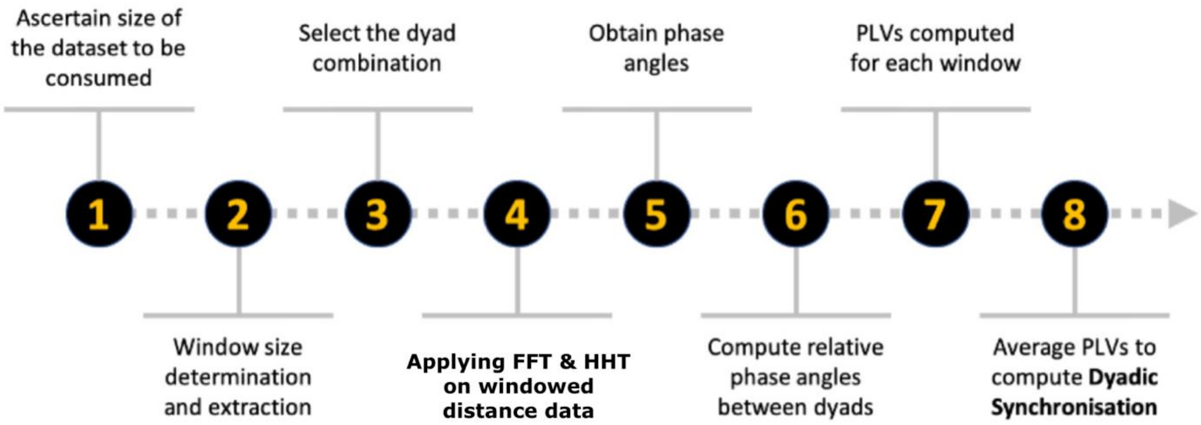

Ascertain size of the dataset to be consumed
Select the dyad combination
Obtain phase angles
PLVs computed for each window
1
2
3
4
5
6
7
8
Window size determination and extraction
Applying FFT & HHT on windowed distance data
Compute relative phase angles between dyads
Average PLVs to compute Dyadic Synchronisation

Figure 2. An illustration of the process pipeline for computing Dyadic Synchronization between a dyad, or in other words two co-performers.

References:

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

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