Physiological Synchrony Revealed by Delayed Coincidence Count: Application to a Cooperative Complex Environment

Kevin J. Verdière, Mélisande Albert, Frédéric Dehais, and Raphaëlle N. Roy

***Abstract*—Synchrony at the physiological level is an objective measure that can be used to investigate cooperation between human agents. This physiological synchrony has been experimentally observed in different dyadic contexts through measures of the autonomous system such as cardiac measures. Various metrics are used to characterize synchrony between participants such as cross-correlation, weighted coherence, or cross recurrence quantification analysis and with a wide variety of paradigms. We propose the delayed coincidence count as a new method for assessing cardiac synchrony. Delayed coincidence count has already been used to characterize synchrony in firing neurons populations. While being straightforward and computationally light, this method has already been formally proven to be statistically robust. A complex dynamic microworld is designed with two difficulty levels and two cooperation conditions. A total of 40 participants, i.e., 20 teams, voluntarily has conducted the experiment. The delayed coincidence count method (with a coincidence threshold δ of 20 ms) reveals a significant synchrony ($p < .01$) during the cooperative and high difficulty condition only, while the other methods did not. The results are interpreted in terms of interaction intensity in accordance with recent literature.**

***Index Terms*—Cooperation, delayed coincidence count, dyad, electrocardiogram (ECG), physiological synchrony.**

I. Introduction

COOPERATION is what allowed living organisms to evolve from multicellular organisms to social insects to reach our current human society [1]. Humans are now by far engaged in the most complex systems of cooperation among living individuals [1]. A general definition of cooperation could be stated as: "a situation that contains a manifest collective goal, in which a group of agents realize it by choosing their actions in accordance with an equilibrium" [2]. While the study of cooperation has been for long the realm of psychosociology and subjective measures, several studies have attempted to identify objective correlates of teammate's synchrony. This field of research known as "interpersonal physiology" or "physiological synchrony" (PS) aims at assessing temporal similarity in teammates' physiological responses [3]. Objectively characterizing synchrony require the acquisition of several data streams from teammates [4] such as electrodermal activity, thermal activity, respiration or cardiac activity [3], [5]. So far, the latter has been the most popular technique to uncover PS in various dyadic contexts, i.e., experiments involving a pair of participants, such as parent-child, couples, therapist-client or teammates (see [3] for a systematic review).

To assess PS, several methods have been applied to the measure of concurrent cardiac signals using cross-correlation, weighted coherence, or cross recurrence quantification analysis (CRQA) [3]. Cross-correlation is a measure of similarity that is computed by a sliding dot product of two different signals. The weighted coherence, introduced by Porges *et al.* [6], is a measure derived from coherence. Coherence can be seen as a correlation coefficient in the frequency domain to characterize how much two signals oscillate in the same frequency band. The weighted coherence uses the frequency power of each of the two signals to weight each frequency bins. Finally, recurrence analysis allows to observe complex and sometimes subtle oscillatory time series behaviors. The rationale for recurrence analysis is that any "time series describing a high-dimensional system composed of multiple coupled variables can be reconstructed from but a single measured variable of that system" [7]. The method of time delays allows us to reconstruct systems in higher dimensions. Once the data are reconstructed in a higher dimension space via time delay, a distance matrix between all possible points can be computed. Each point in this matrix represents the distance between two points of the signals. Points spaced by less than a threshold distance will be considered recurrent. The threshold distance matrix is the recurrence plot (basis of the recurrence analysis). Recurrent quantification analysis (RQA) intends to quantify this dynamics. Cross recurrence uses the same principle to identify the complex oscillatory dynamics of two systems via two signals and in the same way, CRQA uses the same methods as RQA to quantify this dynamics [7].

Physiological synchrony has been shown to be predictive of team performance (i.e., task completion time) using weighted coherence on heart rate measures [8]. However, the authors

This work of Kevin J. Verdière was supported by a PhD grant delivered by the Direction Générale de l'Armement (DGA). The work of Raphaëlle N. Roy and Mélisande Albert was supported by the Neurotools project funded by ISAE-SUPAERO and INSA Toulouse. This article was recommended by Associate Editor D. Wu. *(Corresponding author: Kevin Verdière.)*

Kevin J. Verdière, Frédéric Dehais, and Raphaëlle N. Roy are with the ISAE-SUPAERO, University of Toulouse, 31062 Toulouse, France (e-mail: kevin.verdiere@isae-supaero.fr; frederic.dehais@isae.fr; raphaelle.roy@isae.fr).

Mélisande Albert is with the INSA Toulouse, University of Toulouse, 31062 Toulouse, France (e-mail: melisande.albert@insa-toulouse.fr).

Digital Object Identifier 10.1109/THMS.2020.2986417

pointed out that there was a "lack of a predictive relationship between physiological synchrony and the team coordination," where "coordination" was measured via cross-correlation on their physical joystick action. Similarly, Montague *et al.* [9] evaluated PS with 24 teams with shared experience involving active and passive users. Using weighted coherence on cardiac interbeat interval (IBI), they showed that synchrony relates to group performance after controlling for task/technology and is also correlated with shared perceptions of trust in technology among group members. Jarvela *et al.* [10] studied physiological synchrony across 41 teams playing video games in cooperative or competitive conditions and with or without allies. Using a similar approach to [9], they demonstrated that physiological synchrony correlates with reported empathy between players. More interestingly, they show that the competitive configuration without allies leads to more synchrony, raising the idea that during competition without allies "the players automatically focus more on each other which might turn the game more competitive also experientially." Their viewpoint is that to understand the opposite player, players simulate their behavior and responses within themselves, which is consequently reflected in their body through similar reactions observable via physiological signals synchrony.

Chanel *et al.* [11] studied 21 teams playing a video game in cooperative versus competitive configurations via correlation and weighted coherence on their IBI. Their study revealed that PS increased with subjective players' involvement in the social interaction with higher PS for competitive versus cooperative game. They theorized that PS might be an index of the intensity of interactional behavior and could be used to measure social presence. Elkins *et al.* [12] studied 10 teams of 4 during a military building cleaning task. They showed that a higher physiological synchrony was associated with better team performance and concluded that PS seems to be a part of proficiency in real-world military settings. Conversely, Strang *et al.* [13] did not find any increase in physiological synchrony during cooperative behavior. They used cross-correlation, cross-fuzzy entropy, and CRQA to quantify physiological synchrony while 80 participants played a cooperative Tetris.

Despite these studies, interpersonal autonomous synchrony is still an underexplored research area [3]. As stated by Ekman *et al.* [14] in their review, there is "a general lack of knowledge on how structural elements of the social situation are reflected in psychophysiology." Indeed, these studies rely on different measures and protocols, thus, preventing to draw comparisons and conclusions regarding the underlying physiological mechanisms of PS [3]. One possible approach to better understand PS is to assess it at the heart beat level per se. Fundamental electrophysiological studies characterize neuronal synchrony via spike coincidence analysis [15]. Similarly, one could apply such a method to measure how much two hearts do beat together. Technically, this method relies on the delayed coincidence count metric [16]. The delayed coincidence count represents, in a given range of time for two distinct electrocardiogram (ECGs), the number of beats that occur at the same time, i.e., that are coincident. One advantage of this method is that it allows to account for local phenomena when the coherence and cross-correlation approaches are less sensitive to temporally local variations. Moreover, its implementation and its physiological interpretation are far more straightforward, especially compared to CRQA.

Thus, this study proposes to evaluate the ability of a theoretically robust and yet computationally simple PS method: a permutation method based on a delayed coincidence count to detect heart synchrony during cooperation using a piloting-like task, the multiattribute task battery (MATBII) [17]. We demonstrate its usefulness in assessing cooperation between teammates who perform the task in various difficulties and cooperation settings. Finally, we compare this new method with the most currently used metrics [3], namely, the cross-correlation, the weighted coherence, and the CRQA. Regarding the CRQA, the most used measures are as follows: recurrence rate, determinism, entropy, and average length [3], [7]. First in the Materials and Methods section, the proposed cardiac synchrony metric and the most used ones are described, as well as the experimental protocol used to test them. Next, the subjective and behavioral results concerning the task accomplishment are reported, followed by the synchrony results. Finally, they are discussed with respect to the literature.

II. Materials and Methods

A. Participants

In total, 40 participants (i.e., 20 teams; 9 females; 27 years old ± 8) voluntarily underwent this experiment. They were recruited among the students of the ISAE-SUPAERO Engineering School, Toulouse, France. Of 40, 38 of them were directly recruited as dyads, the remaining 2 were arbitrarily assigned to each other. Out of the 20 dyads, only 2 were not same-sex (i.e., 15 male-male, 3 female-female, 2 female-male). As verified through a questionnaire, 11 considered their teammate as a friend, 6 as an acquaintance, 1 as a family member, 1 as a stranger, and 1 as a lover. All had normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. The study was approved by the local ethic committee (IRB number: IRB00011835-2019-05-28-129) and all participants gave their informed written consent.

B. Experimental Design

1) NASA MATBII and Difficulty Level: A modified version of the MATBII initially developed by NASA was used [17]. The MATBII is "a computer based task designed to evaluate operator performance and workload" [17]. The original version is freely available on the NASA website [18].

As shown in Fig. 1, it is composed of four subtasks, a system/alarm monitoring task (SYSMON), a tracking task (TRACK), a fuel/resource management task (RESMAN), and a communication task (COMM). The system monitoring task requires the participants to respond as quickly as possible to lights and scale fluctuations via keystrokes (F1 to F6). The tracking task requires the participants to keep the circle as close to the center as possible using a joystick. The resource monitoring task requires them to keep the tank A and tank B levels as close to 2500 as possible via managing pumps 1–8

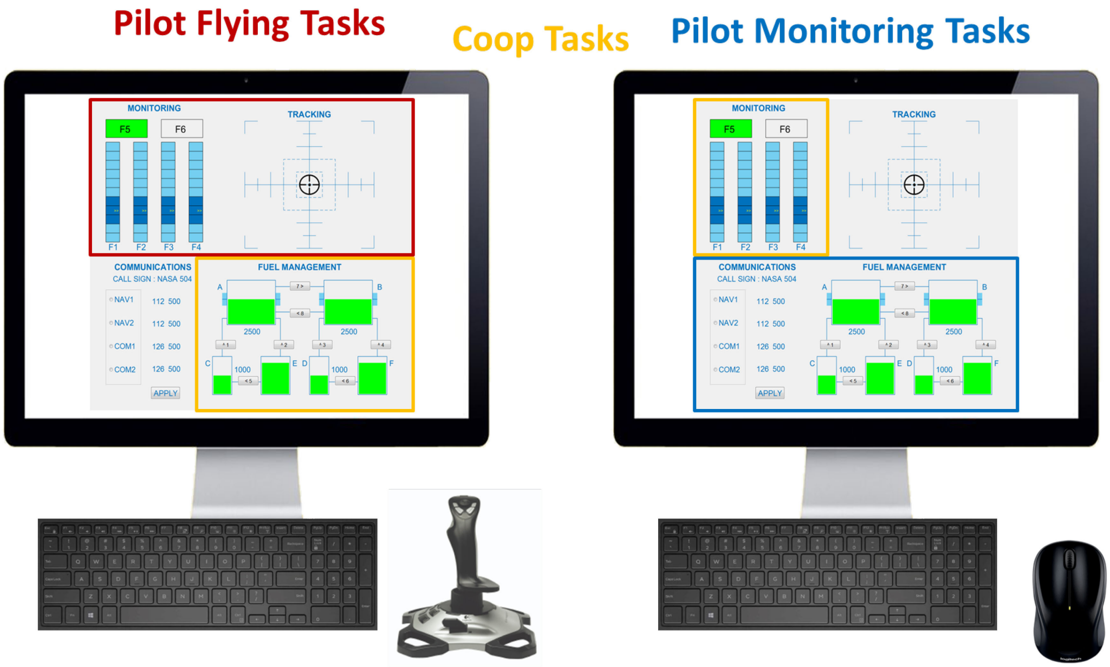
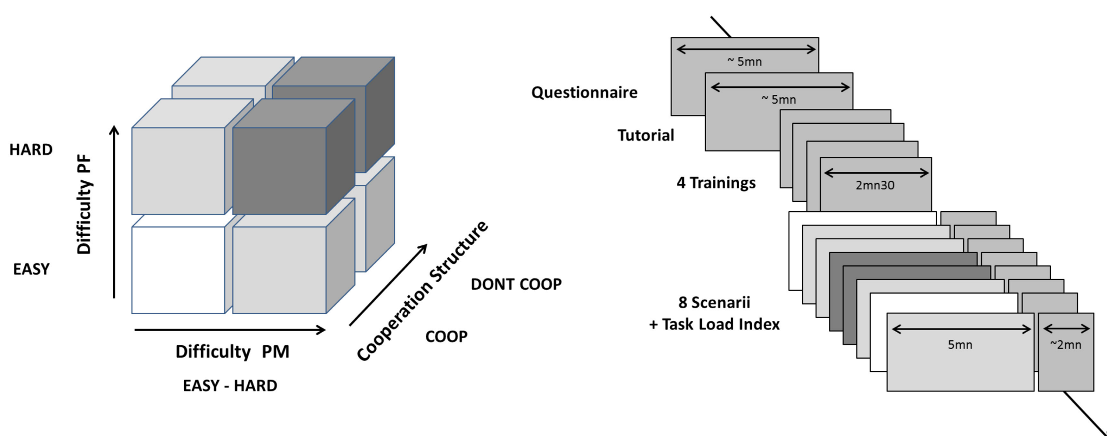


Fig. 1. Modified version of the MATBII. Participants were seated side by side in front of duplicated screens represented here on the left for the pilot flying and on the right for the pilot monitoring. The pilot flying in red had to perform the two upper tasks: monitoring and tracking. The pilot monitoring in blue had to perform the two lower tasks: fuel management and communications. During the cooperative condition, they both had to monitor one of each other's tasks and help to perform it if needed: The pilot monitoring had to monitor and help for the monitoring task and the pilot flying the fuel management task.



(a) (b)

Fig. 2. (a) Graphical representation of the 2^3 factorial design. The three axes represent the three experimental factors. The pilot monitoring difficulty (PM), the pilot flying difficulty (PF), and, finally, the cooperation structure. (b) Experimental timeline.

with the keyboard or the mouse. Finally, the communication task requires the participants to answer to broadcast messages to their call name by indicating the radio and the number heard.

Participants were seated side by side in front of duplicated screens (figure 1). Participant 1 on the left side was called "pilot flying" and had to perform the two upper tasks, namely, the SYSMON and TRACK tasks. He/she had a keyboard and a joystick to do so. Participant 2, called pilot monitoring, had to perform the two lower tasks: RESMAN and COMM. He/she had a keyboard and a mouse to do so. The task difficulty for the pilot flying and pilot monitoring were modulated independently. They were modulated only by changing the difficulty of the TRACK and RESMAN tasks. The number of alarms (SYSMON) and communications (COMM) during each scenario remained the same. There were two levels of difficulty: EASY and HARD. As the difficulty of the task was modulated independently for each teammate, it gave rise to four different difficulty conditions (EASY-EASY, EASY-HARD, HARD-EASY, and HARD-HARD) where the left and right represent the difficulty, respectively, for the pilot flying and the pilot monitoring (see Fig. 2).

2) Cooperation Level: Each participant was attributed two subtasks, however, in order to induce cooperation between the teammates, in half the experimental blocks, the participants had to cross-monitor their partner, i.e., COOP condition.

Cross-monitoring means that participants had to help their partner when possible without speaking. For example, when the participants are in the COOP condition, if the pilot monitoring who is supposed to do the RESMAN and COMM tasks sees that an alarm is ON on the SYSMON task, he/she can respond to it with his/her keyboard in order to improve the overall performance. Hence, in the control condition: DONT COOP, they did not have to cross-monitor each other, but had to do their own two tasks: the two upper and two lower tasks for the pilot flying and pilot monitoring, respectively. Whereas in the COOP condition, they had to perform their own tasks and to cross-monitor the RESMAN and SYSMON for the pilot flying and pilot monitoring, respectively (see Fig. 1).

Moreover, a dependency between the TRACK and RESMAN tasks was implemented in both cooperative and noncooperative scenarios in order to make the two participants environment dependent and, therefore, more realistic [19]. When the tracker was outside the biggest square, it became red and all the pumps of the RESMAN task were deactivated until the tracker came back inside the biggest square. As the tracker was controlled by the pilot flying, this had an influence on the pilot monitoring which managed the RESMAN task. Conversely, when the tank A or tank B levels were under 2000 or above 3000, the TRACK task responsiveness decreased, making it more difficult. Hence, pilot's monitoring actions had an influence on the pilot flying TRACK task.

3) Scenarii and Protocol: Combining the cooperation level (i.e., cross-monitoring: COOP and control condition: DON'T COOP) with the four difficulty combinations, there was a total of eight different scenarii. Each scenario was presented once to the participants in a 5-min block each and in a random order (see Fig. 2).

Once arrived, participants were randomly attributed one role: either pilot flying or pilot monitoring and were asked to seat down. They were asked to fill an informed consent and a demographic questionnaire. Once done, they were given the written task instructions. While they were reading, ECG electrodes were put in place. Participants were able to ask questions regarding the task if needed. Before starting the task, they were asked to seat as comfortably as possible. They were seated approximately 1 m from each other, as in a cockpit. They did a short interactive tutorial, which gave them the occasion to discover and interact with each subtask separately. This tutorial was followed by four training sessions of 2.5 min each. Each training was set in the control condition (DON'T COOP) meaning they did not have to cross-monitor their partner. The first one was an EASY-EASY and the second one a HARD-HARD scenario. The third and fourth were the same but they had to exchange role, the pilot flying did the pilot monitoring job and the pilot monitoring did the pilot flying job. This was done in order to train each participant to do all the tasks so they could help their partner during the cooperation condition if needed. Participants were asked to do their best in order to achieve the best performance. Out of the 20 teams, the best performing one won a flight in a Vulcanair P68 twin engine aircraft in order to motivate the students.

C. Data Acquisition and Analysis

All the analyses were done using MATLAB r2019a. Codes to compute the delayed coincidence count and the permutation test are freely available on github [20].

1) Subjective Assessment: After each scenario, participants were asked to fill a commonly used workload questionnaire: the NASA-TLX [21]. This questionnaire combines six factors, i.e., mental demand, physical demand, temporal demand, overall performance, frustration level, and effort.

2) Behavioral Data:

a) Performance: Performance was rated out of 400 for each scenario (100 for each task). The SYSMON task was evaluated using the average response time, 0 being 7 s and 100 being 0.5 s. The TRACK task was evaluated as the average distance from the center, 0 being the border and 100 the center. The RESMAN task was evaluated as the average distance from 2500 units, 0 being 1000 and 100 being 0. The COMM task was evaluated as the number of good answers: 10 being 100 and 0 being 0.

b) Cooperation: Cooperation was evaluated via participants' keystrokes. It was considered that the pilot flying cooperated when he/she helped by modulating the activity in the RESMAN task, i.e., activating or deactivating a pump (by pressing a number from 1 to 8). Regarding the pilot monitoring, it was considered that he/she helped when he/she responded to alarms of the SYSMON task (i.e., pressing a number from F1 to F6 when needed). A percentage was then computed representing the number of keystroke performed by the Helper over the total number of keystroke performed for this tasks. For example, for the SYSMON subtasks alarms, if the pilot monitoring responded to 3 alarms out of the 30, the percentage would be 10%. The pilot flying would have then responded to the 27 other alarms.

3) ECG Data: ECG was recorded with two BioSemi Active2 (Corp) at 512 Hz. Two electrodes were used, placed under the right clavicle and the left mid-axillary line. The overall ECG pipeline is detailed hereafter in Fig. 4. First raw signals were band pass filtered between 1 and 30 Hz, using a Butterworth filter of the fifth-order. Signals were then epoched to separate the eight different 5 mn scenarios. Peak detection was performed automatically using the "findpeaks" MATLAB function using two parameters. The first parameter is a minimum peak amplitude or height. This means that to be considered as a peak, the value must be above a threshold V_{th}. The default value for V_{th} was set to half the signal maximum value. The second parameter used was a minimum interpeak distance. This parameter can be seen as a refractory period and the default value was set to 250 ms. A visual inspection was performed to dismiss low quality recordings. There were mainly due to movement artifact and electrodes coming off. If needed, the two parameters were manually adjusted. Data were then stored as a time vector containing each peak appearance. From the time vector, the average beat per minute (BPM) was computed. The standard deviation of all NN interval (SDNN) was also computed, where NN interval represent all the "normal" RR interval. A $2 \times 2 \times 2$ repeated

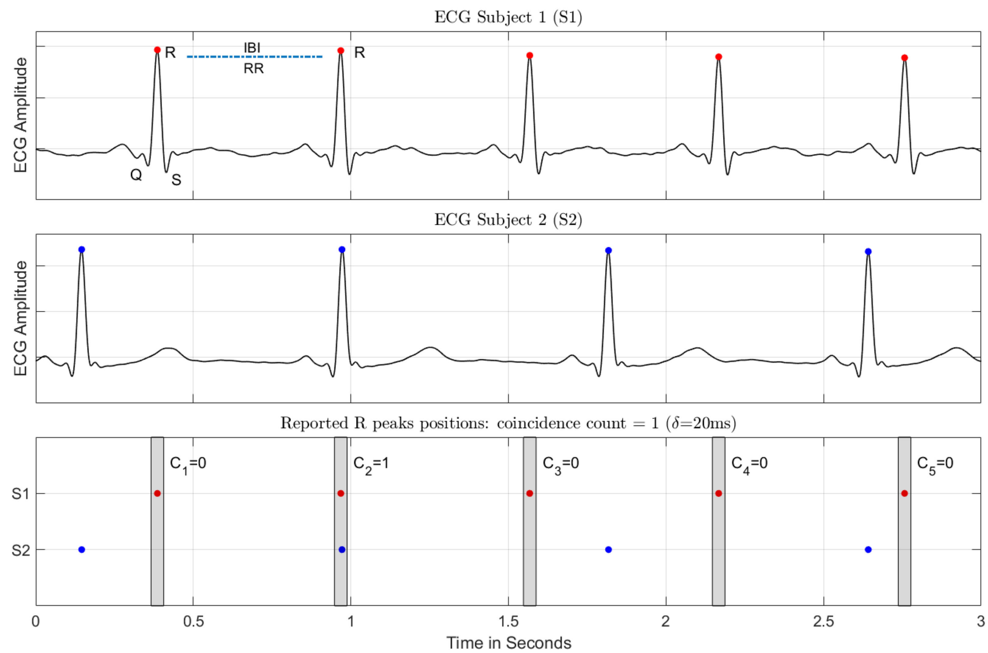


Fig. 3. Illustration of the delayed coincidence count. Two 3-s ECG recording are depicted in the upper and middle graphs (Participants 1 and 2). The red and blue dots represent the ECG R peaks for the first (S1) and second (S2) participants, respectively. On the upper graph, the letter QRS symbolize the first QRS complex. The blue dash line represents the IBI also known as RR interval regarding the R peaks. RR interval can also been called NN interval for "normal" beats. The red and blue ECG peak dots are reported on the lower graph. Coincidence count for this segment is represented here. The first red dot on the left has not blue dot within a time range of $\delta = 20$ ms from it; The count for this first point is then $C_1 = 0$. Conversely, the second red dot has a count $C_2 = 1$ because he was one blue point within a 20 ms range; meaning that the two participants R peaks are coincident. The total coincident count C^t for this segment would be $C^t = \sum_{n=1}^{5} C_n$.

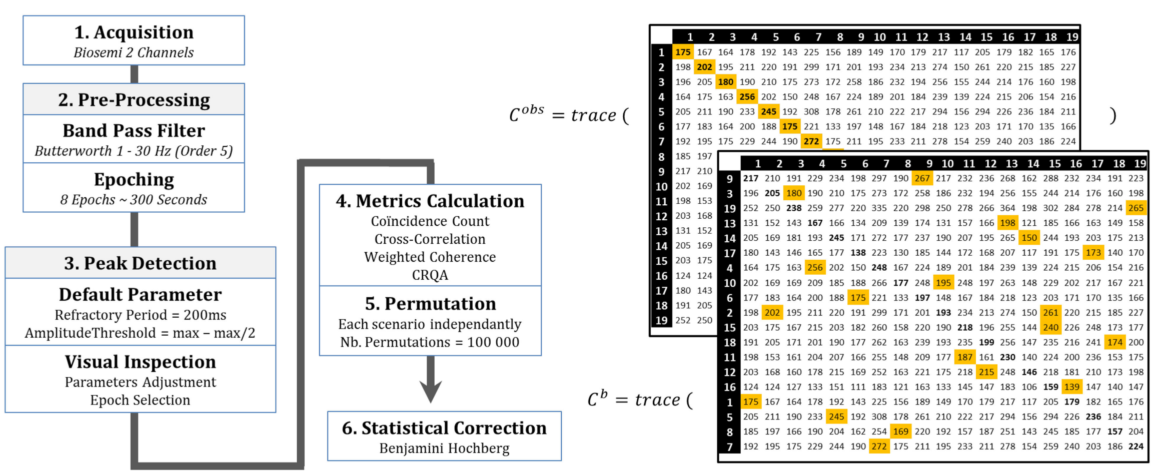


Fig. 4. Left: description of the processing pipeline. Right: two coincidence count matrices, where lines represent pilots flying and columns pilots monitoring. On the original matrix in the back, the diagonal represents the coincidence count for actual couples (highlighted in yellow). C^{Obs} is the diagonal sum of this coincidence diagonal (trace). In front is represented one possible permutation of the original matrix, where lines are shuffled. The diagonal represents now a random association of couples and C^b is the trace of this matrix. Matrices are 19×19 (instead of 20×20) because one couple was excluded for this scenario due to insufficient ECG data quality.

measures analysis of variance (ANOVA) was done for the BPM and SDNN data with the pilot flying difficulty (EASY/HARD), pilot monitoring difficulty (EASY/HARD), and cooperation structure (COOP/DONT COOP) as factors.

D. Cardiac Synchrony Measures

1) Cross-Correlation, Coherence, and Cross Recurrence: The pipeline used for cross-correlation and coherence is similar to the one in Jrvel *et al.'s* [10] study. Cross-correlations at zero lag were obtained through standard procedures. Regarding the weighted coherence [22], it was computed for the frequency ranging from 0.05 to 1.25 Hz using 256 point Hann windows with 75% overlap, weighted by both participants series power spectral values at the specified frequencies.

CRQA was done following the procedure in [23] using the cross recurrence plot toolbox for MATLAB [24]. The four recurrence measures [7] used here were: (a) recurrence rate: the number of shared locations in the phase space, which represents how often systems do synchronize. (b) determinism: quantifying the number of points belonging to a diagonal line and representing how much systems do stay in a synchronized state. (c) Entropy: the diagonal lengths distributed over an histogram and qualifying the system complexity. Finally, (d) the length: the longest diagonal segment in the recurrence plot which describes the chaoticity of the system (see [7] for an comprehensive presentation).

2) Delayed Coincidence Count: In order to evaluate how many ECG peaks were coincident between the pilot flying and the pilot monitoring, we used the delayed coincidence count (as defined in [25]). The delayed coincidence count between two point processes X^1 and X^2 is given by

$$\varphi_\delta^{\text{coinc}}(X^1, X^2) = \sum_{u \in X^1} \sum_{v \in X^2} 1_{|u-v| \leq \delta}. \tag{1}$$

More informally, $\varphi_\delta^{\text{coinc}}$ is the number of couples of spikes (peaks) appearing with a delay at most equal to δ. The two point processes studied here were the pilot flying and pilot monitoring R peaks (X^{PF} and X^{PM}, respectively). We calculated this coincidence count $\varphi_\delta^{\text{coinc}}$ for each team and also between all the pilot flying and pilot monitoring from other team

$$a_{i,j} = \varphi_\delta^{\text{coinc}}(X_i^{\text{PF}}, X_j^{\text{PM}}) \tag{2}$$

for all (i,j) in $\{1, \ldots, n\}^2$ where X_i^{PF} (respectively, X_j^{PM}) represents the pilot's flying (respectively the pilot's monitoring) ECG peaks from the ith (respectively, the jth) couple, and n=20. This means that, for example, for the pilot flying for couple 1 (X_1^{PF}), we calculated the coincidence between him/her and all the other pilots monitoring from couple 1 to 20 (X_j^{PM} with j in $\{1, \ldots, 20\}$). As each scenario was processed independently, if some data were missing for one pilot monitoring or pilot flying, the whole team was excluded for the scenario. By doing so, we obtained eight square coincidence matrices $a_{i,j}$, i.e., one per scenario (see Fig. 4).

The process to set and select δ is similar to the radius selection for recurrence plot as described by Webber Jr. and Zbilut [7]. The δ parameter is fixed based on two notions: 1) δ has to be as small as possible. A large value of δ would make the coincidence count really high, since it would consider the coincidence of each beat with all the other beats. Additionally, the meaning of coincidence itself in the context of heart beats would not be relevant for values above a second (heart rates close to 1/s). 2) δ should not be too small, indeed very small values of δ would drastically reduce the coincidence count, or even zero it. Moreover, very small δ values would also increase the standard deviation of the coincidence count while it has to be considered normalized by its mean. Indeed, as δ increases the coincident count inexorably increases. The coefficient of variation (i.e., the standard deviation divided by the mean) is used to quantify this phenomenon. Hence, in order to select a suitable value for δ, the coincidence count is computed for a range of δ and the optimal value is then selected by choosing the one that minimizes both the coefficient of variation and the value of δ itself. In this study, the δ value was selected at the group level, i.e., regarding all participants' data.

a) Permutation Test: Permutation testing is a nonparametric method to statistically test for samples differences. The idea is to shuffle the data to estimate the sampling distribution and then to compare it to the "real data." Teammates that did the experiment together, i.e., "real teams" are here the "real data." The different teams are supposed to be independent. By shuffling those teams, i.e., creating "permutated teams," we computed the coincidence distribution under the null hypothesis (no synchrony). An example of "real team" could be team number 3: X_3^{PF} and X_3^{PM}. Conversely, a permuted team represents a random association of a pilot flying and a pilot monitoring: X_2^{PF} and X_7^{PM}, for example. This permutation method allowed to compare the number of coincidences of "real teams" from the one of "permuted teams." The coincidence number C^{obs} for "real teams" is computed for each scenario independently. It corresponds to the diagonal sum also known as the trace of the coincidence matrix $a_{i,j}$

$$C^{obs} = \sum_i a_{i,i}. \tag{3}$$

The permutation step consists of drawing B independent and identically distributed permutations $\prod_n^b, 1 \leq b \leq B$ and computing C^b

$$C^b = \sum_i a_{i, \prod_n^b(i)}. \tag{4}$$

This permutation step can be seen as a shuffling between teams illustrated in Fig. 4. It consists of a random pilot flying (X^{PF}) to a random pilot monitoring (X^{PM}) association and then computing the coincidence sum between all those new permuted teams. The computational way to see this permutation is a shuffling between lines of the coincidence matrix $a_{i,j}$. By shuffling lines, a random association between a pilot flying and a pilot monitoring (X^{PF} and X^{PM}) is done on the diagonal. The trace (sum of the diagonal) of this permuted matrix equals C^b. To statistically detect a cardiac significant, the sum of coincidence count for the "real teams" C^{obs} must be significantly higher than the one on randomly permutated teammates (which recreate what happens under independence). To evaluate this, the *p* value *p* is evaluated as follows:

$$p = \frac{1}{B+1}\left(1 + \sum_{b=1}^{B} 1_{C^b \geq C^{obs}}\right). \tag{5}$$

As eight scenarii were evaluated, a false discovery rate (FDR) detection was applied on the *p* value [26].

III. Results

A. Subjective and Behavioral Data

First, in order to validate that the two implemented difficulty conditions (EASY and HARD) were perceived as such, the subjective ratings from the NASA TLX questionnaire were compared (see Fig. 5). The pilots flying found the task significantly more difficult when the condition was HARD than when it was EASY ($F(1, 19) = 68.25$, $p < 10^{-3}$, $\eta_p^2 = .78$). The pilots monitoring also found the task significantly more difficult when the condition was HARD ($F(1, 19) = 41.17$, $p < 10^{-3}$, $\eta_p^2 = .68$). Interestingly, the pilots monitoring also found the task significantly more difficult in the COOP condition than in the DON'T-COOP condition ($F(1, 19) = 5.46, p < .05$,

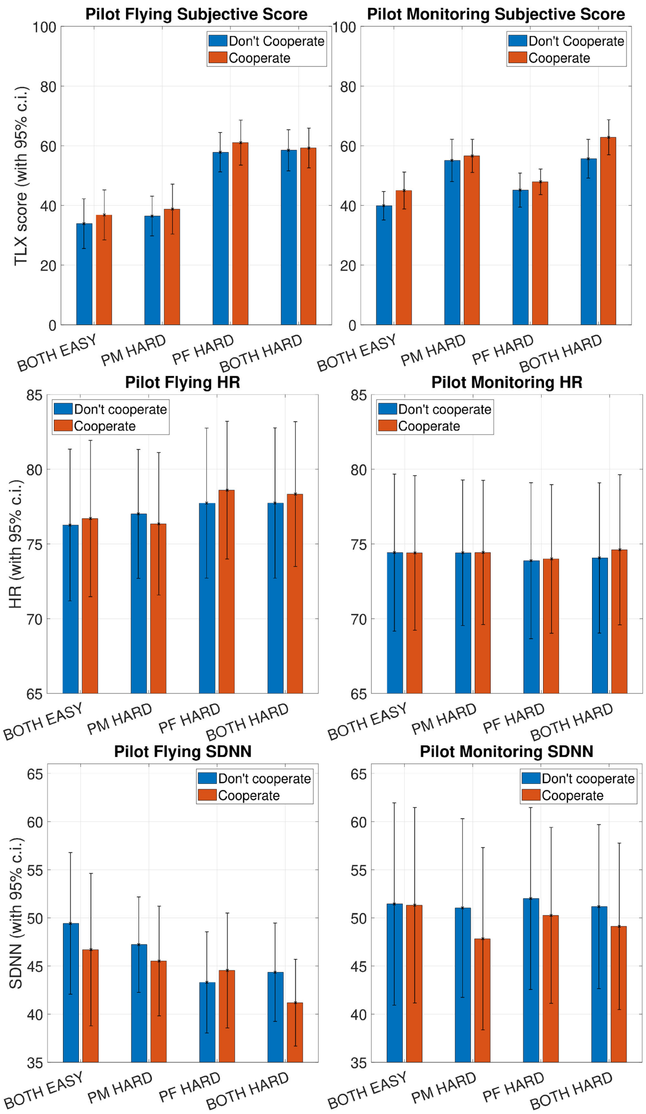


Fig. 5. Subjective NASA TLX scores (first row), average heart rate (second row), and SDNN (third row) for the pilots flying (left graph) and the pilots monitoring (right graph). Each one of the eight bars represents one of the eight scenarii. The x-axis corresponds to the scenario difficulty for the pilot flying and the pilot monitoring (i.e., EASY-EASY/EASY-HARD/HARD-EASY/HARD-HARD). Colors represent the cooperation condition (COOP–DON'T COOP).

$\eta_p^2 = .22$). The pilots flying did not significantly found the task more difficult in the COOP condition ($F(1, 19) = 3.57, p = .07$, $\eta_p^2 = .16$).

Next, as expected, the implemented task difficulty had an effect on the overall performance (see Fig. 7). Note that the performance is evaluated for the team as the whole and not for each teammate individually. Teammates exhibited a significantly higher performance when the pilot flying was in the EASY condition compared to when he/she was in the HARD condition ($F(1, 19) = 16.61$, $p < 10^{-3}$, $\eta_p^2 = .46$). Similarly, teammates also performed better when the pilot monitoring was in the EASY condition ($F(1, 19) = 11.84$, $p < .01$, $\eta_p^2 = .38$). Interestingly, the COOP condition also had a significant effect on performance: participants performed slightly better in the NO-COOP condition than in the COOP condition ($F(1, 19) = 9.19$, $p < .01$, $\eta_p^2 = .33$).

Regarding the cooperation, the keystroke percentage was only evaluated for the four scenarios where teammates were asked to cooperate (COOP). This keystroke percentage represents quantitatively how much a teammate did help his/her partner. It corresponds to the number of keystrokes performed by a teammate in the other teammate's subtasks compared to the total number of keystrokes performed in this subtasks. As expected, the pilots flying cooperated less ($F(1, 19) = 5.83, p < .05$, and $\eta_p^2 = .23$) when their difficulty was HARD ($M = 8.9, std = 13.2$) compared to when it was EASY ($M = 13.7, std = 14.9$). In the same way, the pilots monitoring cooperated less ($F(1, 19) = 8.71$, $p < .01$, and $\eta_p^2 = .80$) when their difficulty was HARD ($M = 11.6, std = 13.4$) compared to when it was easy ($M = 11.1, std = 15.1$).

B. ECG Data

After visual inspection, some scenarii were dismissed due to nonsufficient ECG data quality. When one portion of a scenario had to be dismissed, instead of interpolating the missing part, the whole scenario was dismissed for this participant. This unfortunate loss of data is mainly due to the fact that we used external electrodes of the Biosemi system that encountered loose contact issues for the ground and reference electrodes. In the end, 81% and 95% of scenarii were retained, respectively, for pilots flying and pilots monitoring.

Regarding the heartrate (HR), the average pilot flying HR was significantly higher ($F(1, 16) = 6.23$, $p < .05$, and $\eta_p^2 = .28$) when their difficulty was HARD ($M = 78.5, std = 10.7$) compared to when it was EASY ($M = 76.6, std = 11.2$). Neither the pilot monitoring difficulty nor the cooperation condition had a significant effect on the pilot flying HR. Regarding their teammates (i.e., pilot monitoring), difficulty or the cooperation conditions had no significant effect on the HR.

Regarding heartrate variability (HRV) measures, for the pilot flying, it appears that the SDNN was significantly higher ($F(1, 16) = 9.68$, $p < .01$, $\eta_p^2 = .38$) when it was EASY for him/her ($M = 47.2, std = 15.1$) compared to when it was HARD ($M = 43.0, std = 10.1$). Surprisingly, the difficulty of the pilot monitoring tasks had also an effect ($F(1, 16) = 6.92$, $p < .05$, and $\eta_p^2 = .30$). The pilot flying SDNN was significantly higher when it was EASY for the pilot monitoring ($M = 46.1, std = 13.5$) compared to when it was HARD ($M = 43.8, Std = 13.6$). The cooperation condition did not exhibit a significant effect on the pilot flying SDNN. Concerning the pilot monitoring SDNN, neither their difficulty, the pilot flying difficulty, nor the cooperation condition had a significant effect.

C. Cardiac Synchrony

1) Cross-Correlation, Weighted Coherence, and CRQA: Cross-correlation at zero-lag and weighted coherence metrics revealed no significant synchrony via the permutation test—p values were above the corrected threshold. Regarding CRQA,

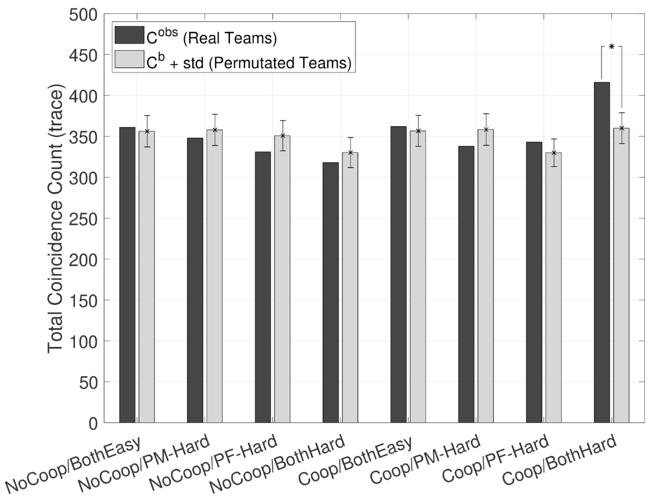


Fig. 6. Coincidence matrices traces for real teams (C^{obs}) and 100 000 permutated teams (C^{b}) for the eight scenarii ($\delta = 20$ ms). The standard deviation is represented only for permutated teams (C^{b}) since only one value per scenario exists for C^{Obs}. The permutation test revealed a significant difference ($p < 0.01$) for the eighth scenario (COOP–HARD–HARD).

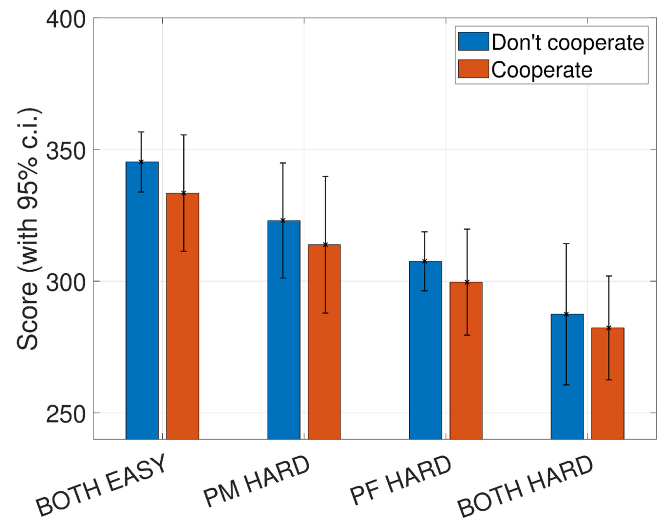


Fig. 7. Team task performance. The eight bars each represent one of the eight scenarii. The x-axis corresponds to the difficulty of the scenario for the pilot flying—pilot monitoring. (EASY–EASY/EASY–HARD/HARD–EASY/HARD–HARD). Colors represent the cooperation condition (COOP–DON'T COOP). Values range from 0 to 400, 400 being a perfect score.

the data were normalized and the parameters were set following the procedure and recommendations described in [7]. The used parameters were: $M = 4$ for the embedded dimension, $\tau = 1$ for the delay, and $r = .1$ for the radius. The four metrics that were used are recurrence rate, determinism, length, and entropy. As detailed in Section II-D, to statistically assess the synchrony, the permutation test was done for the eight scenarii independently. As eight tests were performed, an FDR correction was applied on the p value. It revealed no significant synchrony for those four metrics. All p values were above the corrected threshold.

2) Delayed Coincidence Count: The optimal threshold limit δ parametrized to compute the coincidence count was 20 ms. The statistical permutation test procedure was exactly the same as the one for cross-correlation, coherence, and CRQA. The value for the total coincidence count, i.e., the trace of the $a_{i,j}$ matrices are represented in Fig. 6. The total coincidence count for the 100 000 permutations, i.e., the $a_{i,j}$ matrices with lines shuffled are represented alongside them. Interestingly, the eighth scenario, which corresponds to both teammates operating in a difficult condition (i.e., HARD–HARD) and in a cooperation condition (COOP) revealed a significant cardiac synchrony between teammates ($p < .01$) for a maximum time delay of 20 ms. Note that all the other scenarii did not elicit such a cardiac synchrony. Additionally, no significant correlation was found between the coincidence count and the performance index across cooperative conditions.

IV. Discussion

The goal of this study was to evaluate PS during dyadic interactions using a delayed coincidence detection method applied on ECG R peaks. The main interests of this method are its ease of implementation and its ability to account for local cardiac synchrony. The method proved efficient in characterizing physiological synchrony in dyads that performed a highly engaging piloting-like task in a cooperative setting. Thus, 20 teams had to perform a dual multi-attribute task battery MATBII task in which the levels of difficulty and cooperation were manipulated. This method was then compared with the most used metrics in the literature: cross-correlation, weighted coherence, and CRQA.

The subjective and behavioral findings confirmed the task to be engaging and contrasted in terms of workload. Indeed, teammates performed better and reported a lower mental effort when facing the easy conditions than the hard ones. Moreover, task difficulty modulated the ability to cooperate. In this task, the cooperation condition required the pilot flying and pilot monitoring to crosscheck their partner's actions and user interface and to potentially assist them. Our behavioral and subjective results disclosed that this was particularly challenging under demanding settings (HARD–HARD–COOP) yielding the participants to be more focused on handling their own task and leaving them less time and cognitive resources to assist each other. Cooperation also intrinsically increases the number of tasks to perform, and, therefore, the operator's workload. Hence, the obtained result is consistent with previous cooperative studies indicating the mental workload had a deleterious effect on cross-checking and crew performance [27].

Interestingly enough, this latter demanding condition was the only one to elicit significant PS as calculated by the delayed coincidence count method. On the one hand, one could argue that this effect could be explained in terms of higher HR for the teammates induced by the HARD–HARD conditions, thus artificially increasing the heart beat coincidence count. However, this effect was not observed in the HARD–HARD–DON'T COOP condition. Moreover, only pilots flying had a significant heart rate increase during their HARD difficulty. On the other hand, our results did not lead to observe PS in any of the other cooperative situations (e.g., EASY–EASY–COOP). Therefore, we believe that our results account for both the workload and the intensity of cooperation that occurred in the HARD–HARD–COOP condition. The participants were particularly engaged in performing their own task while having in mind that they had to support each other. This conclusion is akin to that of Levenson and Gottman [28] who made the connection with results on marital interaction, and to Chanel *et al.*'s study who reported a greater physiological synchrony during conflict interaction compared to low-conflict discussion [11]. In their study, they observed more PS in a competitive versus a cooperative condition while playing a video game. They described PS as a "candidate for interaction intensity." These results are, however, to be qualified. Indeed, the coincidence counts were not correlated to the performance index. This is contradictory to only part of the literature that found that PS was predictive of team performance in some aspect [8], [10]–[12]. Yet is not in-line either to the other studies that report no increase in PS during cooperative behavior [13]. Our study does highlight an increase in PS during cooperative and high workload conditions, without correlation with team performance. This might reveal that the observed physiological synchrony could be an epiphenomenon.

The results reported in this study, together with others [8], [10]–[12], [23], [29]–[31], raise the issue of the mechanisms that underlie cardiac synchronization between teammates. Researchers proposed different theories regarding the source such as "shared metabolic demand through matched activity or behavior, conditional and environmental influences and synchronized breathing" [3]. Spontaneous group synchrony has been observed via breathing [32]. This phenomenon known as the chameleon effect [33] was also highlighted during cooperative conversation [34] or visual and verbal interaction [35]. Respiratory coupling or more generally breathing might play a role in the observed synchrony. Moreover, the task design itself can induce short stress episodes linked to the dynamic and fluctuating workload experienced by participants. Those short episodes can be linked to breathing synchrony, which could result in cardiac synchrony.

In addition, it should be noted that the task was designed to be continuous, in opposition to turn-based tasks, and to engage the two participants during both the easy and hard conditions. Keeping participants active in the task was done via the continuous nature of the tracking and resource management task. Moreover, because of the implemented dependencies between the subtasks, the encountered workload has been variable for each team for an exact same difficulty. This is mainly noticeable when one of the participants performed poorly, the strong dependency between the tracking and resource management tasks increase drastically the difficulty for the coparticipant. This particularly explains why the difficulty of the pilot monitoring tasks impacted the pilot flying's physiological state, such as her/his SDNN. In other words, the overall task difficulty was controlled and equal between teams, but we only looked at the performance at the team level opposed to the participant level.

Regarding data analysis, most of the previous studies used correlation or cross-correlation, weighted coherence, or recurrence analysis, i.e., CRQA on the IBI. Coherence and correlation metrics did not reveal any PS. This might be due first to the fact that scenarii were too short as they lasted only 300 s (i.e., 5 min), which might not be optimal for computing those metrics. Most importantly, we can hypothesize that because of the task difficulty, cooperative behavior arises only sporadically. Thereby, methods such as coherence and correlation might not be appropriate as they characterize an average linkage throughout time. Hence, they could be thought of measuring temporally global synchronies, contrary to the coincidence detection metric, which measures temporally local synchronies.

The closest method to compare ours to seems to be CRQA. CRQA estimates the dependencies between each point of two signals in a reconstructed higher dimension space. The dependencies are estimated via the thresholded point distances in this reconstructed space, i.e., the recurrence plot. By doing so, it can characterize oscillatory behavior and complex dynamics such as nonlinear coupling and chaotic behavior. Theoretically, our method operates really closely by computing distances between points of two signals and counting the number of distances below a fixed threshold. However, the two methods differ regarding the signal used. Our method measures distances using ECG peak appearance time values, whereas CRQA is based on the IBI values. Moreover, CRQA uses a reconstructed dimension space to compute distances, whereas our method directly computes distances in time. Yet, surprisingly the CRQA metrics did not reveal any synchrony. This might be mainly due to the fact that CRQA uses the IBI values. As the IBI is by definition the interval between ECG peaks, it indirectly represents participants' HR. A reconstructed space with time-delayed dimension would then exhibit close points where both participants' HR would vary similarly. In this context, CRQA would detect synchronous HR variations rather than synchronous heartbeats. For this reason, CRQA might not be the most suitable method for characterizing PS in our ecological context because of the nature of the considered coupling itself. Prolonged cooperative behavior might be observable in a recurrence plot based on IBI, the short duration (5 min) of our scenarii could also explain this result. However, we can also hypothesize that because of the "sporadic" nature of cooperative behavior, CRQA might not be the most suitable metric to characterize it in ecological conditions.

To conclude, this study indicated that a highly difficult task combined with a cooperative behavior induces a cardiac synchrony, which can be assessed using a permutation test on a delayed coincidence count. This result is interesting for two main reasons: 1) cooperation states can be measured via cardiac synchrony; and 2) this synchrony can be easily characterized from computational and theoretical points of view. This study brings a contribution to this overall objective of characterizing the level where information appears to transfer between people that cooperate. For the future, we believe that a systematic experimental approach is still needed to evaluate, extract, and isolate every possible source of synchrony between participants. Hence, research improvements such as verifying the impact of the location of the teammate, of the number of noncritical subtasks and their resulting workload, as well as the impact of breathing on cardiac synchrony should be pursued.

References

[1] M. A. Nowak, "Five rules for the evolution of cooperation," *Sci.*, vol. 314, no. 5805, pp. 1560–1563, 2006.

[2] C. Paternotte, "Minimal cooperation," *Philosophy. Social Sci.*, vol. 44, no. 1, pp. 45–73, 2014.

[3] R. V. Palumbo *et al.*, "Interpersonal autonomic physiology: A systematic review of the literature," *Personality Social Psychol. Rev.*, vol. 21, no. 2, pp. 99–141, 2017.

[4] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen, "Interpersonal synchrony: A survey of evaluation methods across disciplines," *IEEE Trans. Affect. Comput.*, vol. 3, no. 3, pp. 349–365, Jul.–Sep. 2012.

[5] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, *Handbook of Psychophysiology*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[6] S. W. Porges, R. E. Bohrer, M. N. Cheung, F. Drasgow, P. M. McCabe, and G. Keren, "New time-series statistic for detecting rhythmic co-occurrence in the frequency domain: The weighted coherence and its application to psychophysiological research," *Psychological Bull.*, vol. 88, no. 3, pp. 580–587, 1980.

[7] C. L. Webber Jr and J. P. Zbilut, "Recurrence quantification analysis of nonlinear dynamical systems," *Tut. Contemporary Nonlinear Methods Behavioral Sci.*, vol. 94, pp. 26–94, 2005.

[8] R. A. Henning, W. Boucsein, and M. C. Gil, "Social–physiological compliance as a determinant of team performance," *Int. J. Psychophysiology*, vol. 40, no. 3, pp. 221–232, 2001.

[9] E. Montague, J. Xu, and E. Chiou, "Shared experiences of technology and trust: An experimental study of physiological compliance between active and passive users in technology-mediated collaborative encounters," *IEEE Trans. Human-Machine Syst.*, vol. 44, no. 5, pp. 614–624, Oct. 2014.

[10] S. Järvelä, J. M. Kivikangas, J. Ktsyri, and N. Ravaja, "Physiological linkage of dyadic gaming experience," *Simul. Gaming*, vol. 45, no. 1, pp. 24–40, 2014.

[11] G. Chanel, J. M. Kivikangas, and N. Ravaja, "Physiological compliance for social gaming analysis: Cooperative versus competitive play," *Interacting Comput.*, vol. 24, no. 4, pp. 306–316, 2012.

[12] A. N. Elkins, E. R. Muth, A. W. Hoover, A. D. Walker, T. L. Carpenter, and F. S. Switzer, "Physiological compliance and team performance," *Appl. Ergonom.*, vol. 40, no. 6, pp. 997–1003, 2009.

[13] A. J. Strang, G. J. Funke, S. M. Russell, A. W. Dukes, and M. S. Middendorf, "Physio-behavioral coupling in a cooperative team task: contributors and relations," *J. Exp. Psychol.: Human Perception Perform.*, vol. 40, no. 1, pp. 145–158, 2014.

[14] I. Ekman, G. Chanel, S. Järvelä, J. M. Kivikangas, M. Salminen, and N. Ravaja, "Social interaction in games: Measuring physiological linkage and social presence," *Simul. Gaming*, vol. 43, no. 3, pp. 321–338, 2012.

[15] P. Uhlhaas *et al.*, "Neural synchrony in cortical networks: history, concept and current status," *Frontiers Integrative Neuroscience*, vol. 3, 2009, Art. no. 17.

[16] S. Grün, M. Diesmann, and A. Aertsen, "Unitary events in multiple single-neuron spiking activity: I. detection and significance," *Neural Comput.*, vol. 14, no. 1, pp. 43–80, 2002.

[17] Y. Santiago-Espada, R. Myer, K. Latorella, and J. Comstock, "The multi-attribute task battery ii (MATB-II) software for human performance and workload research: A users guide (NASA/TM-2011–217164)," NASA Langley Research Center, Hampton, VA, USA, 2011.

[18] NASA (Producer), "Revised multi attribute task battery official website," 2016. [Online]. Available: https://matb.larc.nasa.gov

[19] T. Liu and M. Pelowski, "Clarifying the interaction types in two-person neuroscience research," *Frontiers Human Neurosc.*, vol. 8, 2014, Art. no. 276.

[20] K. Verdière, "Delayed coincidence count and permutation methods," 2019. [Online]. Available https://github.com/

[21] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in *Adv. Psychol*, vol. 52, pp. 139–183, 1988.

[22] V. K. Yeragani, A. Barua, I. Sattiraju, and N. Mallavarapu, "Weighted coherence: A more effective measure than average coherence," *Cardiovascular Eng.: An Int. J.*, vol. 3, no. 4, pp. 141–146, 2003.

[23] K. Kodama, S. Tanaka, D. Shimizu, K. Hori, and H. Matsui, "Heart rate synchrony in psychological counseling: A case study," *Psychol.*, vol. 9, no. 7, pp. 1858–1874, 2018.

[24] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Phys. Rep.*, vol. 438, nos. 5/6, pp. 237–329, 2007.

[25] C. Tuleau-Malot, A. Rouis, F. Grammont, and P. Reynaud-Bouret, "Multiple tests based on a Gaussian approximation of the unitary events method with delayed coincidence count," *Neural Comput.*, vol. 26, no. 7, pp. 1408–1454, 2014.

[26] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc.: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[27] F. Dehais, J. Behrend, V. Peysakhovich, M. Causse, and C. D. Wickens, "Pilot flying and pilot monitorings aircraft state awareness during go-around execution in aviation: A behavioral and eye tracking study," *The Int. J. Aerosp. Psychol.*, vol. 27, nos. 1/2, pp. 15–28, 2017.

[28] R. W. Levenson and J. M. Gottman, "Marital interaction: Physiological linkage and affective exchange," *J. Personality Social Psychol.*, vol. 45, no. 3, pp. 587–597, 1983.

[29] S. J. Wilson *et al.*, "When couples hearts beat together: Synchrony in heart rate variability during conflict predicts heightened inflammation throughout the day," *Psychoneuroendocrinology*, vol. 93, pp. 107–116, 2018.

[30] E. Ferrer and J. L. Helm, "Dynamical systems modeling of physiological coregulation in dyadic interactions," *Int. J. Psychophysiology*, vol. 88, no. 3, pp. 296–308, 2013.

[31] R. A. Henning and K. T. Korbelak, "Social-psychophysiological compliance as a predictor of future team performance," *Psychologia*, vol. 48, no. 2, pp. 84–92, 2005.

[32] E. Codrons, N. F. Bernardi, M. Vandoni, and L. Bernardi, "Spontaneous group synchronization of movements and respiratory rhythms," *PLoS One*, vol. 9, no. 9, 2014, Art. no. e107538.

[33] T. L. Chartrand and J. A. Bargh, "The chameleon effect: the perception–behavior link and social interaction," *J. Personality Social Psychol.*, vol. 76, no. 6, pp. 893–910, 1999.

[34] K. Shockley, M.-V. Santana, and C. A. Fowler, "Mutual interpersonal postural constraints are involved in cooperative conversation," *J. Exp. Psychol.: Human Perception Perform.*, vol. 29, no. 2, pp. 326–332, 2003.

[35] M. J. Richardson, K. L. Marsh, and R. Schmidt, "Effects of visual and verbal interaction on unintentional interpersonal coordination," *J. Exp. Psychol.: Human Perception Perform.*, vol. 31, no. 1, pp. 62–79, 2005.