# Unsupervised Clustering of Team Situation Awareness Via Cross Recurrence Quantification Analysis of Psychophysiological Metrics

**Ching-Yu Cheng[1], Liuchuan Yu[2], Lap-Fai Yu[3], and Behzad Esmaeili[4]**

Edwardson School of Industrial Engineering, Purdue University, West Lafayette, IN, USA
Department of Computer Science, George Mason University, Fairfax, VA, USA
cheng640@purdue.edu, lyu20@gmu.edu, craigyu@gmu.edu, besmaei@purdue.edu

**Abstract**

**Team Situation Awareness (TSA), a collective understanding of changes in the surroundings, is critical in human-human and human-robot teaming in the futuristic construction site. Current measuring techniques fall short in capturing continuous and real-time data in realistic settings. Psychophysiological signals, including visual attention (eye-tracking) and neural response (prefrontal cortex activation), have the potential to address this problem. This study investigated the relationship between TSA scores and Cross Recurrence Quantification Analysis (CRQA) metrics derived from psychophysiological data using unsupervised clustering techniques. A two-worker-robot collaboration VR experiment was conducted, in which both participants needed to do a highly interdependent pipe task with each other and cobots, while potential hazards occurred in their proximity. Participants' psychophysiological metrics were collected (42 individuals), and TSA was assessed via the Situation Awareness Global Assessment Technique (SAGAT) survey. CRQA was used to quantify the temporal coupling of pairs, while unsupervised clustering was adopted to discover the latent structure in the coupling. The results demonstrated that CRQA metrics derived from both eye-tracking and prefrontal cortex (PFC) activation serve as valid indicators to classify TSA groups. Specifically, high TSA teams recur their visual attention more frequently, more simultaneously, and for longer durations. Plus, high TSA teams share more frequent, more varied, and longer coupling patterns, and have longer alignment of each other's current neural state. The findings serve as proof of concept for using continuous psychophysiological metrics as indicators for TSA, and suggest future research to examine other latent structures between TSA and psychophysiological metrics.**

**Keywords –**

**Team Situation Awareness; Human-Robot Collaboration; Unsupervised Clustering; CRQA; Functional Near-infrared Spectroscopy (fNIRS); Eye-tracking**

## 1 Introduction

Construction workers face significant cognitive challenges in their worksites due to the dynamic nature of construction activities [1]. The construction workers must collaborate with other workers to achieve their production goals, while working in proximity to moving equipment (e.g., trucks, cranes) and changing work environment (e.g., uncovered openings on the floor). Such a work environment has made construction one of the most hazardous industries with a disproportionate number of fatalities and injuries annually [2]. The situation is becoming even more challenging, considering that workers need to collaborate with emerging cobots more frequently, as cobots, such as drones and robot dogs, are being invented and taking over some workers' roles in the near future. To support effective human-robot and human-human collaborations, it is imperative to understand workers' awareness of their tasks, teammates, and the surrounding environment. One of the concepts that helps with this understanding and quantifies human workers' awareness in a multi-agent teamwork setting is team situation awareness (TSA).
Situation awareness (SA) is a person's perception, understanding, and projection of the changes in the surroundings [5], whereas TSA is a collective SA in a team. Psychophysiological measures serve as external manifestations of internal cognitive processes, providing a continuous option to measure TSA. This research utilizes Cross Recurrence Quantification Analysis (CRQA) to explore the latent structure of multiple time-varying signals. However, there is limited knowledge

about the relationship between psychophysiological CRQA metrics and TSA score, and how to use CRQA metrics to assess TSA. To address this knowledge gap, this study examines the association between psychophysiological CRQA metrics and TSA using unsupervised clustering techniques. The results of this study advance our fundamental knowledge regarding the relationship between psychophysiological coordination patterns with TSA and can be used to develop algorithms to practically track TSA and notify workers who might get involved in accidents.

## 2 Background

### 2.1 Measuring Team Situation Awareness

TSA has been studied in nuclear power operations and the military (e.g., [3], [4]). One of the widely used measuring techniques is Situation Awareness Global Assessment Technique (SAGAT), designed by Endsley [5]. Yet, its biggest disadvantage is freezing during simulation, making it inapplicable to the real world [6]. One of the approaches to mitigate this limitation is using psychophysiological sensors, which provide continuous signals [7]. Nevertheless, analyzing multivariate data from multiple team members is complicated, making it difficult to interpret interpersonal correlation. This highlights the need for an analytical method capable of quantifying dynamic structure between paired signals.

### 2.2 Cross Recurrence Quantification Analysis

Analyzing the interactions between two time series is challenging, particularly when the signals are nonlinear and asymmetric, which is common for psychophysiological data collected from different individuals. Cross Recurrence Plot (CRP) and CRQA are two of the tools to address this challenge. CRP visualizes dependency and interrelationships between two nonlinear time series. It maps two signals onto a matrix and marks the recurrence when both revisit the same state (Figure 1). Yet, CRP relies on qualitative visual inspection, which limits objective comparison.

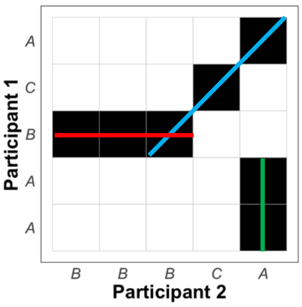



Figure 1. Illustration of the recurrence of letters in the sequence "AABCA" (participant 1) and "BBBCA" (participant 2). The black squares indicate the recurrence of a letter, while white squares describe the absence of recurrence. The red, blue, and green lines represent the recurrence in the direction of horizontal, diagonal, and vertical, respectively.

To complement subjective interpretation, CRQA provides a quantitative method to assess coupling dynamics. It quantifies persistence, diversity, and structure in directions of diagonal, horizontal, and vertical when coupling occurs, providing a bivariate perspective to study dual signals. Previous literature has identified 13 common metrics, as shown in Table 1 [9]. Specifically, the recurrence Rate (RR) indicates the density of recurrence. Other metrics can be roughly categorized based on the directions along diagonal, vertical, and horizontal of the matrix: (1) Determinism/Laminarity explains the ratio of recurrence forming diagonal/vertical/horizontal lines; (2) Maximum length; (3) Average length; and (4) Entropy refers to the complexity of line distribution.

Since the data in this research are from two individuals (i.e., asymmetric), it is valuable to investigate both vertical and horizontal directions. For example, vertical metrics explain the recurrence structure of participant 1 when participant 2 is stable, and horizontal means the reverse. CRQA has been applied to psychophysiological signals, including behavioral data such as eye-tracking [8], and continuous data such as vectorcardiogram [9], electromyography [10], electroencephalogram [11], and fNIRS [12].

### 2.3 Team Situation Awareness in Construction

Previous SA research in construction focused on the individual level, particularly workers' workload and risk perception [13] and limited knowledge is available about the SA among crews. One of the challenges in studying TSA in construction is the multi-demand construction activities, where workers need to coordinate interdependent tasks with teammates and monitor safety hazards in the surroundings simultaneously. In this situation, a single aggregated score for TSA from each aspect of awareness may simplify the underlying cognitive processes and fail to capture how different aspects interact. For example, a worker may be overwhelmed by the task progress and ignore a teammate's status. Such imbalanced awareness cannot be reflected by an overall TSA score. These limitations highlight the need for including multiple dimensions for measuring TSA in a construction environment.

To address this issue, a recent study by the authors decomposed TSA into three dimensions: environment TSA, task TSA, and team TSA [14]. This multidimensional approach represents key focuses of safety, task productivity and quality, and team members' intentions and actions. The decomposition provides a more nuanced insight into which parts of TSA are more influential, and how to address corresponding deficiencies.

Table 1. CRQA Metrics

| | Metrics | Definition in CRP | Interpretations |
|---|---|---|---|
| | Recurrence Rate (RR) | The ratio of recurrence points out of the total points | Higher RR implies higher recurring states. |
| **Diagonal** | Determinism (DET) | The ratio of recurrence points forming diagonal lines | Higher when participants' recurrence is more synchronized, reflecting a more predictable pattern. |
| | Maximum diagonal line length ($L_{max}$) | The length of the longest diagonal line | The longest duration when both participants share synchronized periods. |
| | Average diagonal line length (L) | The average length of diagonal line | Higher when participants have longer average durations of synchronized recurrence. |
| | Entropy (ENTR) | Shannon entropy of the distribution of diagonal line lengths. | Higher when participants' recurrence is distributed more diversely. |
| **Vertical** | Vertical laminarity ($LAM_v$) | The ratio of recurrence points forming vertical lines | Higher when participant 1 has more continuous recurrence, reflecting participant 1's laminar behavior. |
| | Maximum vertical line length ($V_{max}$) | The length of the longest vertical line | The longest duration during which participant 1 has a continuous recurrence. |
| | Vertical trapping time ($TT_v$) | The average length of vertical line | Higher when participant 1 has longer average durations of continuous recurrence. |
| | Vertical entropy ($ENTR_v$) | The entrophy of the distribution of vertical line lengths | Higher when participant 1's recurrence is distributed more diversely. |
| **Horizontal** | Horizontal laminarity ($LAM_h$) | The ratio of recurrence points forming horizontal lines | Higher when participant 2 has more continuous recurrence, reflecting participant 2's laminar behavior. |
| | Maximum horizontal line length ($H_{max}$) | The length of the longest horizontal line | The longest duration during which participant 2 has a continuous recurrence. |
| | Horizontal trapping time ($TT_h$) | The average length of horizontal line | Higher when participant 2 has longer average durations of continuous recurrence. |
| | Horizontal entropy ($ENTR_h$) | The entrophy of the distribution of horizontal line lengths | Higher when participant 2's recurrence is distributed more diversely. |

Note: Assume $R_{i,j}$ indicates the state of participant 1 at time i is compared to the state of participant 2 at time j.

## 3 Point Of Departure

Despite the growing importance of TSA in the construction industry, there is limited knowledge about how to bridge multiple physiological data using CRQA metrics with TSA. This study examines the relationship between TSA and psychophysiological data using CRQA metrics. To achieve this goal, two participants were recruited to do a VR experiment simulating realistic tasks and hazards, and their psychophysiological data and SAGAT scores were recorded. To explore interpersonal visual and neural couplings, this study selects eye-tracking and PFC activation, capturing complementary aspects of teamwork. These data were analyzed using both CRQA metrics to quantify coupling dynamics and unsupervised clustering techniques to explore latent coordination patterns.

Based on this framework, the hypotheses are:

H1. CRQA metrics produced from eye-tracking differ significantly between high/low TSA groups.

H2. CRQA metrics produced from PFC activation differ significantly between high/low TSA groups.

## 4 Methodology

### 4.1 VR Experiment design

A two-worker-robot collaboration VR experiment was designed to simulate real-world dynamics of a construction site [15]. The experimental setup required participants to coordinate and cooperate with their teammates, including both another human participant and cobots such as a robot dog and a drone, continuously during safety events in the surroundings and their shared pipe installation tasks in a 10 min span. This section is structured based on three TSA dimensions: environment, task, and team, from 4.1.1 to 4.1.3.

#### 4.1.1 Environment

Common construction hazards, such as struck-by and falling objects, are simulated in the VR through virtual construction vehicles such as a truck, a drone, and a tower crane. These construction vehicles randomly work in participants' proximity, requiring the team's continuous vigilance. When vehicles start to move, both participants hear the operating sound and a warning cue (e.g., "Warning, a truck is backing up nearby"). If participants were unaware of vehicles, they may be involved in simulated accidents, such as being struck by a truck or hit by falling pipes. Therefore, both participants need to be actively attentive to their surroundings to sustain the environment TSA (Figure 2).

#### 4.1.2 Task

The collaborative pipe installation task was highly

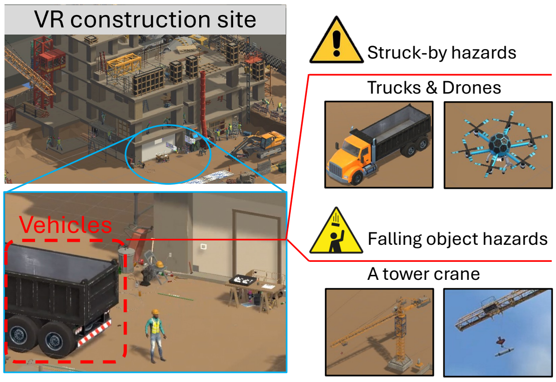


Figure 2. VR Environment and potential hazards

interdependent, requiring continuous teamwork from complementary roles and asymmetric duties. The pipe installation task can be divided into several steps in sequence: exchanging specs, ordering, cutting, and installing the pipe. Player 1 has more task loadings than player 2. Specifically, player 1 ordered/installed pipes and ordered glue and clamps, whereas player 2 cut pipes and ordered connectors. When the experiment started, both participants received partial pipe specs, and they needed to share with each other to get a full picture of the pipe specs. In addition, cobot teammates assisted participants in different ways: a delivery drone carried pipes for player 1 whenever an order was placed by player 1, and a robot dog cut pipes for player 2. This teamwork requires efficient communication and collaboration to achieve task success, thereby fostering task TSA (Figure 3).

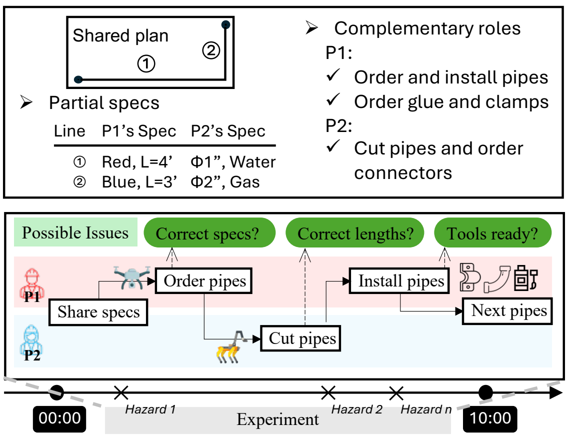


Figure 3. Pipe Task and Workflow

### 4.1.3 Team

Participants were expected to track their human teammate's actions for real-time progress (i.e., productivity) and safety concerns. For example, player 2 relied on player 1 to provide pipe length information before cutting, while both participants depended on timely mutual updates to avoid accidents. They also needed to track their cobot teammate's status for the task progress. The frequent tracking of teammates required the participants' team TSA (Figure 4).

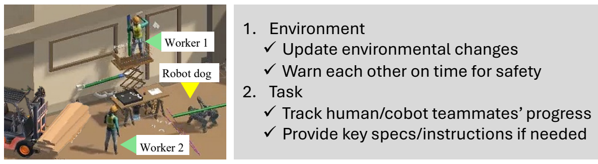


Figure 4. Teamwork to achieve productivity and safety goals

## 4.2 Participants

The research team recruited 21 dyads (42 individuals) from Purdue's campus (age m = 24.17, sd = 4.24) to participate in the experiment. To reduce the learning effect, every participant spent 2.5 hours learning the fundamental skills before the experiment started, including the virtual interface, task procedures, and hazard identification. All participants communicated fluently in English during the experiment.

## 4.3 Apparatus

To collect eye data in the VR settings, the Meta Quest Pro was chosen, featuring powerful eye-tracking and full-body tracking capabilities. To explore PFC activation, the fNIRS device was utilized to study fluctuations in oxygenated and deoxygenated hemoglobin on the scalp.

## 4.4 Data Collection and Preprocessing

Psychophysiological data (including PFC activation and eye-tracking) were collected during the experiment and were extracted only when safety events occurred. Specifically, eight safety events were extracted with durations ranging from 7s~30s (m=16.88s, sd=8.94s) and approximately one min between consecutive events. Both participants independently completed a SAGAT survey [16] after the experiment. The traditional freezing probe was not implemented to avoid interrupting continuous psychophysiological data collection.

### 4.4.1 Eye-Tracking

Participants' visual attention on the area of interest (AOIs) was categorized into three groups, matching TSA's three dimensions. Environment AOIs included dynamic construction vehicles and their dependents. Task AOIs consisted of objects related to the pipe task, such as pipes, connectors, clamps, glue, and the wall. Group AOIs comprised of teammates' avatars and cobots.

### 4.4.2 Prefrontal Cortex Activation

PFC processes executive functions, including decision-making, working memory, and speech [17, 18]. Right and left PFC ($PFC_R$ and $PFC_L$) were the regions of interest (ROIs) in this study. $PFC_R$ was related to spatial working memory, attentional control, and negative affect, while $PFC_L$ was associated with verbal working memory and positive affect [19, 20]. Previous studies have identified increased PFC activation during demanding tasks [21].

fNIRS optodes were placed on Brodman Area 9 and 46: Dorsolateral prefrontal cortex (DLPFC), which is sensitive to cognitive workload. The layout included two 10-long channels and two short channels (Figure 5).

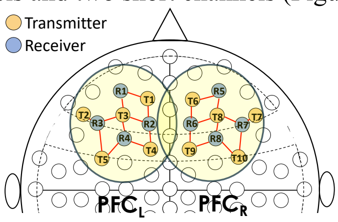


Figure 5. fNIRS optode layout

### 4.4.3 TSA Score

SAGAT queries were structured into 3 dimensions (environment, task, and team) and 3 SA levels. An individual's SA score was the aggregation of all correct answers in the SAGAT query, and the TSA score was the summation of both player 1 and player 2.

### 4.4.4 Cross Recurrence Quantification Analysis

Before CRQA, the psychophysiological data were visually inspected for completeness, and missing data were removed (e.g., group 1 player 2's eye tracking data). Eye tracking data were analyzed at the dyad-level (n=20), while PFC data were analyzed at the hemisphere-level (n=42).

One of the critical steps in CRQA is to decide on analysis parameters, which were computed following a general CRQA preprocessing pipeline [22], including determining parameters of delay, embedding, and radius (Table 2). To reduce variability in individuals, PFC activation was normalized using z-score within each event. CRQA was then performed on both eye tracking data and PFC activation for eight events, and its metrics were averaged across events for unsupervised clustering.

Table 2. CRQA Parameters

| | Eye-Tracking | PFC Activation |
|---|---|---|
| Characteristics | Categorical | Continuous |
| Delay | 1 | 1 |
| Embedding | 1 | 3 |
| Norm | NA | Euclidean |
| Radius | 0 | 0.29 |

### 4.4.5 Unsupervised Clustering

To reduce dimensionality, Principal Component Analysis (PCA) was utilized on CRQA metrics. Each CRQA metric was normalized across teams before PCA, and the first two PCs were retained for clustering. Two clustering techniques, K-Means clustering and Gaussian Mixture Models (GMM), were employed. Clustering validity was evaluated using the Silhouette score, ranging from -1 to 1, with a value close to 1 indicating good clustering (i.e., cohesion and separation).

Lastly, average TSA scores were compared between clusters. Statistical analyses were used to quantify group differences, including the Mann-Whitney U (MWU) test, permutation test, and Cliff's delta. Cliff's delta serves as an effect size, ranging from -1 to 1, with values farther from zero indicating a stronger separation.

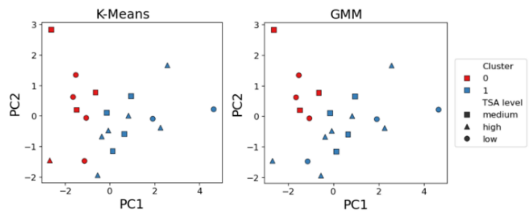
## 5 Results

The results of principal component analysis are shown in Table 3. Eye tracking's first two PCs include RR, the ratio of recurrence forming diagonal, vertical, and horizontal lines (DET/LAM), and average length along these directions (L/TT) (Table 3). PC1 exhibits a general increase in all selected CRQA metrics. A higher PC1 score indicates that both participants look at the same AOIs more often (RR), more simultaneously (DET), and for longer (L), and participant 1 revisited AOIs longer ($TT_v$) and more stably ($LAM_v$). In contrast, PC2 focuses on the asymmetric visual coupling pattern (positive $TT_h$, negative $LAM_v$, and $TT_v$). A higher PC2 indicates that participant 2 dwells longer on AOIs ($TT_h$), and participant 1 frequently changes AOIs ($LAM_v$) and watches for less time ($TT_v$).

Table 3. PC's Explained Variance and Loadings

| Metrics | Eye-Tracking | | PFC Activation | |
|---|---|---|---|---|
| PC's explained variance | 0.631 | | 0.925 | |
| PCs | PC1 | PC2 | PC1 | PC2 |
| RR | 0.529 | 0.109 | 0.391 | 0.003 |
| DET | 0.511 | 0.119 | - | - |
| L | 0.397 | -0.317 | 0.405 | -0.045 |
| ENTR | - | - | 0.391 | -0.054 |
| $LAM_h$ | 0.349 | 0.069 | - | - |
| $TT_h$ | 0.301 | 0.62 | 0.387 | -0.326 |
| $ENTR_h$ | - | - | 0.346 | -0.576 |
| $LAM_v$ | 0.216 | -0.594 | - | - |
| $TT_v$ | 0.206 | -0.363 | 0.382 | 0.385 |
| $ENTR_v$ | - | - | 0.338 | 0.64 |

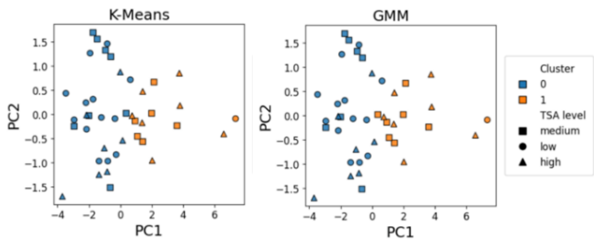
In Figure 6, the groups can be roughly divided into two clusters, visually based on their PC1 values: higher PC1 (cluster 1) and lower PC1 (cluster 0), suggesting that teams in cluster 1 exhibit greater overall visual coupling than those in cluster 0.



(a) K-Means (b) GMM

Figure 6. Scatter Plot in PCA space for Eye-Tracking

PFC Activation's first two PCs include RR, the average length (L/TT), and complexity along directions (ENTR) (Table 3). PC1 has positive loadings on all CRQA metrics. A higher PC1 value reflects that both participants recurred PFC signals more often (RR), with the coupling lasting longer (L) and distributed more varied (ENTR), and participant 2 recurred PFC signals longer ($TT_h$) with more diverse distribution ($ENTR_h$). PC2, on the other hand, again exhibits the unbalanced neural coupling pattern (Negative $TT_h$/$ENTR_h$, positive $TT_v$/$ENTR_v$). A higher PC2 indicates that participant 1 exhibited more varied distribution ($ENTR_v$), whereas participant 2's recurrence is shorter ($TT_h$) and distributed more uniformly ($ENTR_h$). In Figure 7, groups in cluster 1 have a higher PC1 score than those in cluster 0.



(a) K-Means (b) GMM

Figure 7. Scatter Plot in PCA space for PFC Activation

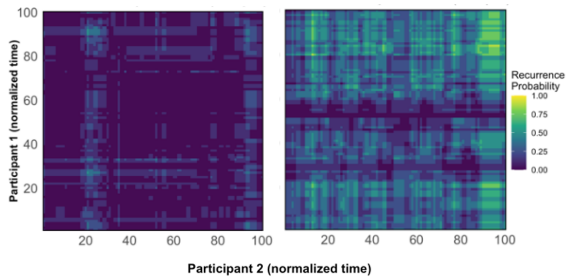
The clustering performance of TSA is summarized in Table 4. For eye-tracking, both K-Means and GMM techniques have silhouette scores greater than 0.27. The average TSA differs significantly (p-values<0.05 for both K-Means and GMM) across clusters, and its Cliff's delta = 0.667. With this statistical support, it can be inferred that high TSA teams tend to align their visual attention more often, more simultaneously, and for longer periods, and participant 1 revisited AOIs longer and more stably than those with poorer TSA. Therefore, H1 is supported.

Table 4. Clustering Performance of TSA Using PCs

| Metrics | Eye-Tracking | | PFC Activation | |
|---|---|---|---|---|
| Clustering | K-Means | GMM | K-Means | GMM |
| Silhouette | 0.315 | 0.272 | 0.483 | 0.471 |
| cluster 0 | n=8 | n=6 | n=28 | n=27 |
| mean TSA | 20.00 | 19.33 | 22.21 | 22.22 |
| cluster 1 | n=12 | n=14 | n=14 | n=15 |
| mean TSA | 25.67 | 25.14 | 25.14 | 24.93 |
| Cliff's delta | 0.667 | 0.667 | 0.352 | 0.321 |
| p $value_1$ | $0.014^{*}$ | $0.022^{**}$ | $0.066^{*}$ | $0.088^{*}$ |
| p $value_2$ | $0.006^{**}$ | $0.010^{**}$ | $0.065^{*}$ | $0.079^{*}$ |

Note: p $value_1$ from the MWU test, p $value_2$ from the permutation test; * < 0.1; ** < 0.05;

In Figure 8, teams in cluster 1 have more frequent and longer diagonal lines (brighter in recurrence probability), reflecting stronger visual coupling. Conversely, teams with higher PC2 are relatively darker, indicating fewer and shorter diagonal lines.



(a) Cluster 0 (b) Cluster 1

Figure 8. Eye-Tracking CRP

For PFC activation, both K-Means and GMM yielded silhouette scores above 0.47, and the average TSA differs marginally significantly (0.05<p-values <0.1 for both K-Means and GMM) across clusters (Table 4). It suggests that high TSA teams couple neural signals more frequently for longer duration, recur more diversely, and participant 2 recurred PFC signals longer with a more diverse distribution. Hence, H2 is partially supported. As shown in Figure 10, both $PFC_L$ and $PFC_R$ CRP in cluster 1 are brighter, distributed more variably diagonally and horizontally, and contain longer diagonal and horizontal lines. Conversely, both $PFC_L$ and $PFC_R$ in Figure 9 are darker, indicating fewer and shorter lines.

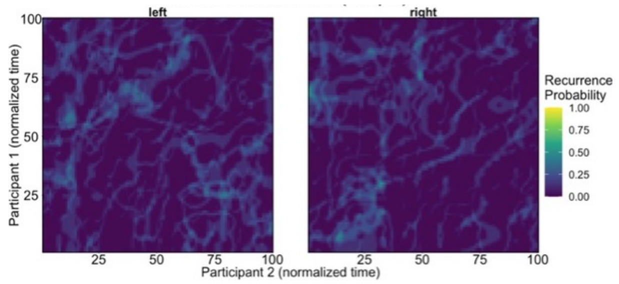


Figure 9. PFC Activation CRP in cluster 0

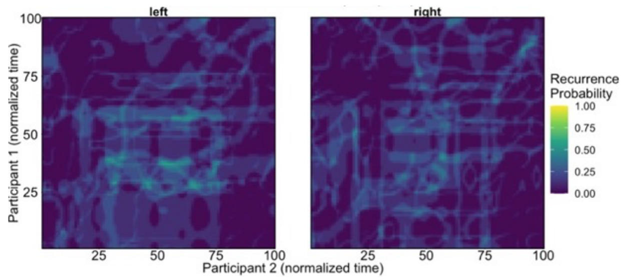


Figure 10. PFC Activation CRP in cluster 1

## 6 Discussions

This study discovers a subtle association between the TSA score and both eye-tracking and PFC activation using CRQA metrics and unsupervised clustering, exhibiting the potential of using continuous measuring techniques to distinguish a team's collective understanding. The findings support the first hypothesis:

High and low TSA teams have significantly different patterns in the visual coupling pattern. Our inference was aligned with the idea that higher gaze recurrence facilitates smoother interaction and better understanding and anticipation of teammates [23, 24], which could be a possible factor of improved TSA. On the contrary, findings from Villamor argued that successful dyadic teams in programming debugging have lower RR, DET, and L, possibly because team members tend to search for bugs independently rather than identical scanpaths [25]. The inconsistency from the above studies may stem from task nature. Further study can explore the relationship between TSA score and eye CRQA metrics under different conditions, such as under time pressure, to better quantify how collective understandings emerge through visual coordination.

Another hypothesis is also supported: High and low TSA teams display marginally significantly different neural coupling patterns. This paper is one of the first to explore the relationship between CRQA-based neural signals and TSA levels. Clusters in both eye-tracking and PFC activation could be roughly divided along PC1 instead of PC2, indicating some high and low TSA teams shared similar characteristics of PC2. This suggests an unbalanced neural coordination pattern rather than uniformly stronger coupling, which aligns with Lu [26]. Furthermore, prior hyperscanning research on shared intentionality [27] revealed that interpersonal neural synchronization emerges when team members share goals and actively cooperate instead of merely coexisting with identical tasks. These distinctive characteristics may be found in high and low TSA teams and explain the latent relationships between TSA scores and neural coupling. The complex teamwork should be conceptualized as a holistic cognitive system, rather than an aggregation of individual cognitive processes, since synchronization in bidirectional interaction is more effective than a unidirectional condition [28].

These findings advance the knowledge of continuous representations of TSA scores, and extend the application beyond discrete and intrusive measures, providing a preliminary reference to explore more structure between the TSA score and multimodal signals. It can be used to support early identification of at-risk teams and provide intervention for them about the upcoming hazards if needed. Furthermore, the results highlight the importance of training workers to distribute their attentions more effectively to foster TSA.

However, this research has some limitations. Firstly, the VR settings may not fully simulate all real-world nuances, such as facial expressions and physics interactions. Secondly, CRQA reveals recurrence instead of the semantic content of attention or neural state [25]. A synthesized analysis of both CRQA and content-specific analysis on each AOI may be a promising direction for future study.

## 7 Conclusions

This study examines the latent relationship between TSA scores and psychophysiological data (both eye-tracking and PFC activation) using CRQA and unsupervised clustering in a dynamic and collaborative VR experiment. The results demonstrated that CRQA metrics derived from psychophysiological responses serve as a valid indicator for categorizing TSA scores into high/low groups. Specifically, team members in high TSA teams recur their attentions more frequently, more simultaneously, and for longer periods than those in low TSA teams. High TSA teams also align their neural states more often and longer, and distribute recurrence more diversely. These findings provide proof of concept for using psychophysiological metrics as continuous and nonintrusive TSA measurements. Future work could incorporate other TSA measuring techniques and analysis methods, such as verbal protocol analysis, to get a more comprehensive picture of TSA theory.

## 8 Acknowledgments

The National Science Foundation is thanked for supporting the research reported in this paper through the Future of Work at the Human-Technology Frontier (FW-HTF) program. This paper was based on work supported by the National Science Foundation under Grant Nos. 2310210 and 2128867. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## 9 References

[1] Alotaibi A, Gambatese J, Tymvios N. Impacts of Work Setting and Task Complexity on Mental Workload and Safety: An Experimental Study. In Construction Research Congress 2024 (pp. 376-387).

[2] U.S. Bureau of Labor Statistics. Census of Fatal Occupational Injuries Summary, 2025 - 2025 A01 Results. (accessed January 2026). https://www.bls.gov/news.release/cfoi.nr0.htm

[3] Mansikka H, Harris D, Virtanen K. Accuracy and similarity of team situation awareness in simulated air combat. Aerospace medicine and human performance. 2023 Jun 1;94(6):429-36.

[4] Takacs VK, Juhasz M. Measuring team situation awareness through team communication: A study on nuclear main control room crews. Journal of Emergency Management (Weston, Mass.). 2024 Mar 1;22(2):139-54.

[5] Endsley MR. Situation awareness global assessment technique (SAGAT). In Proceedings of the IEEE 1988 national aerospace and electronics conference 1988

May 23 (pp. 789-795). IEEE.

[6] Endsley MR, Selcon SJ, Hardiman TD, Croft DG. A comparative analysis of SAGAT and SART for evaluations of situation awareness. In Proceedings of the human factors and ergonomics society annual meeting 1998 Oct (Vol. 42, No. 1, pp. 82-86). Sage CA: Los Angeles, CA: Sage Publications.

[7] Zhang T, Yang J, Liang N, Pitts BJ, Prakah-Asante K, Curry R, Duerstock B, Wachs JP, Yu D. Physiological measurements of situation awareness: a systematic review. Human factors. 2023 Aug;65(5):737-58.

[8] Atweh J, Riggs SL. Quantifying collaborative strategies and identifying performance breakdowns of UAV C2 teams using multidimensional cross-recurrence quantification analysis. International Journal of Human-Computer Studies. 2025 Aug 5:103593.

[9] Meyers A, Buqammaz M, Yang H. Cross-recurrence analysis for pattern matching of multidimensional physiological signals. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2020 Dec 1; 30(12).

[10] Riehm CD, Bonnette S, Rush JL, Diekfuss JA, Koohestani M, Myer GD, Norte GE, Sherman DA. Corticomuscular cross-recurrence analysis reveals between-limb differences in motor control among individuals with ACL reconstruction. Experimental brain research. 2024 Feb;242(2):355-65.

[11] Timothy LT, Krishna BM, Nair U. Classification of mild cognitive impairment EEG using combined recurrence and cross recurrence quantification analysis. International Journal of Psychophysiology. 2017 Oct 1; 120:86-95.

[12] Nguyen T, Hoehl S, Bertenthal BI, Abney DH. Coupling between prefrontal brain activity and respiratory sinus arrhythmia in infants and adults. Developmental cognitive neuroscience. 2022 Dec 1; 58: 101184.

[13] Cheng CY, Esmaeili B. Situation awareness study in the construction industry: A systematic review. In Construction Research Congress 2024 2024 (pp. 843-853).

[14] Cheng, CY., Yu, L., Yu, L., Esmaeili, B. (2025). Visual Allocation of Teams in the Construction Industry: Shared Situation Awareness Under Information Overload in Human-AI Collaboration. AHFE (2025) International Conference, vol 179. AHFE International, USA.

[15] Yu L, Cheng CY, Ranc WF, Dow J, Szilagyi M, Huang H, Hong SR, Esmaeili B, Yu LF. Col-Con: a collaborative and configurable VR platform for construction—a pipe installation case study. 2024

[16] Endsley MR. Situation awareness global assessment technique (SAGAT). In Proceedings of the IEEE 1988 national aerospace and electronics conference 1988 May 23 (pp. 789-795). IEEE.

[17] Baddeley A. Working memory. Memory. 2020 Mar 12:71-111.

[18] Stuss DT, Benson DF. Neuropsychological studies of the frontal lobes. Psychological bulletin. 1984 Jan;95(1):3.

[19] Wu LL, Potenza MN, Zhou N, Kober H, Shi XH, Yip SW, Xu JH, Zhu L, Wang R, Liu GQ, Zhang JT. A role for the right dorsolateral prefrontal cortex in enhancing regulation of both craving and negative emotions in internet gaming disorder: A randomized trial. European Neuropsychopharmacology. 2020 Jul 1;36: 29-37.

[20] Fossati P, Hevenor SJ, Graham SJ, Grady C, Keightley ML, Craik F, Mayberg H. In search of the emotional self: an fMRI study using positive and negative emotional words. American Journal of Psychiatry. 2003 Nov 1; 160(11):1938-45.

[21] Chang WC, Hasanzadeh S. Mental workload in worker-drone communication in future construction: Considering coexistence, cooperation, and collaboration interaction levels. Advanced Engineering Informatics. 2025 May 1;65:103110.

[22] Wallot S, Leonardi G. Analyzing multivariate dynamics using cross-recurrence quantification analysis (crqa), diagonal-cross-recurrence profiles (dcrp), and multidimensional recurrence quantification analysis (mdrqa)–a tutorial in r. Frontiers in psychology. 2018 Dec 4; 9:2232.

[23] Jermann P, Mullins D, Nüssli MA, Dillenbourg P. Collaborative gaze footprints: Correlates of interaction quality.

[24] Uzunosmanoğlu SD, Çakir MP. Examining an online collaboration learning environment with the dual eye-tracking paradigm: The case of virtual math teams. In International Conference on Learning and Collaboration Technologies 2014 Jun 22 (pp. 462-472). Cham: Springer International Publishing.

[25] Villamor M, Rodrigo Mm. Gaze Collaboration Patterns of Successful and Unsuccessful Programming Pairs. In International Conference on Computers in Education 2018 Nov 26.

[26] Lu K, Gao Z, Wang X, Qiao X, He Y, Zhang Y, Hao N. The hyper-brain neural couplings distinguishing high-creative group dynamics: an fNIRS hyperscanning study. Cerebral Cortex. 2023 Mar 1;33(5):1630-42.

[27] Fishburn FA, Murty VP, Hlutkowsky CO, MacGillivray CE, Bemis LM, Murphy ME, Huppert TJ, Perlman SB. Putting our heads together: interpersonal neural synchronization as a biological mechanism for shared intentionality. Social cognitive and affective neuroscience. 2018 Aug;13(8):841-9.

[28] Dai R, Liu R, Liu T, Zhang Z, Xiao X, Sun P, Yu X, Wang D, Zhu C. Holistic cognitive and neural processes: a fNIRS-hyperscanning study on interpersonal sensorimotor synchronization. Social cognitive and affective neuroscience. 2018 Nov;13(11):1141-54.