Identification and Evaluation of Effective Strategies in a Dynamic Visual Task Using Eye Gaze Dynamics

Rakesh Veerabhadrappa (rveerabhadrappa@deakin.edu.au)
Deakin University

Imali T. Hettiarachchi
Deakin University

Samer Hanoun
Deakin University

Dawei Jia
Defence Science and Technology Group

Simon G. Hosking
Defence Science and Technology Group

Asim Bhatti
Deakin University

Research Article

Keywords: unmanned aerial vehicle (UAV), Epistemic network analysis (ENA), Cross recurrence quantification analysis (CRQA), emergency response teams, power station operators, pilots, medical practitioners

Posted Date: August 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-778091/v1

Identification and Evaluation of Effective Strategies in a Dynamic Visual Task Using Eye Gaze Dynamics

Rakesh Veerabhadrappa[1*], Imali T. Hettiarachchi[1], Samer Hanoun[1], Dawei Jia[2], Simon G. Hosking[2], and Asim Bhatti[1*]

[1]Institute for Intelligent Systems Research and Innovation, Deakin University, Waurn Ponds, Australia
[2]Aerospace Division, Defence Science and Technology Group, Melbourne, Australia
[*]rveerabhadrappa@deakin.edu.au
[*]asim.bhatti@deakin.edu.au

ABSTRACT

Simulation-based training utilising visual displays are common in many defence and civil domains. The performance of individuals in these tasks depends on their ability to employ effective visual strategies. Quantifying the performance of the trainees is vitally important when assessing training effectiveness and developing future training requirements. The approach, attitudes and processes of an individual's learning varies from one to another. In this light, some visual strategies may be better suited to the dynamics of a task environment than others, the result of which could be observed in the superior performance outcomes of some individuals. In this study, eye gaze data is used to investigate the relationship between performance outcomes and visual strategies. In an attempt to emulate real operational settings, a challenging task environment using multiple targets that had minimal salient features was selected for the study. Eye gaze of participants performing a simulation-based unmanned aerial vehicle (UAV) refuelling task was used to facilitate the investigation. Cross recurrence quantification analysis (CRQA) and Epistemic network analysis (ENA) were employed on eye gaze data to provide spatial-temporal mapping of visual strategies. A CRQA measure of recurrence rate was used to observe participants' fixation interest on various regions of the task environment. The recurrence behaviours were categorised into cases of visual strategies using an unsupervised clustering algorithm. This article discusses the relationship between the visual strategy cases and performance outcomes to observe which are the most effective. Using the relationship between recurrence rates and performance outcomes, we demonstrate and discuss a gaze-based measure that could objectively quantify performance.

1 Introduction

The effectiveness and efficiency of individuals to achieve task objectives is dependent on their level of proficiency (novice to expert).[1] Knowledge of task-related activities and awareness of informative elements in the task environment greatly influences an individual's task performance.[2] In cognitively challenging domains, professionals (e.g., emergency response teams, power station operators, pilots, and medical practitioners) must interpret complex visual patterns in order to make appropriate decisions. For example, air traffic controllers manage the airspace by comprehending the information available on their radar display to safely guide inbound and outbound aircraft. Students training to become experts in air traffic management are encouraged to develop problem-solving perceptual strategies.[3] Thus, superior performances that are in high demand in cognitively challenging and emergency response domains can be achieved via expert-level perceptual-cognitive skills. These skills enable a thorough understanding of the task structure, comprehension of current events to anticipate future progression, and discernment of any concealed information within work settings.[4–6] To this end, the eye-movement behaviour of individuals has been widely explored as an assessment tool to measure the proficiency level of perceptual-cognitive skills.[6–8]

Studies exploring the relationships between eye-movement behaviour and task performance have extensively employed gaze-based metrics such as saccades and fixation duration,[6] the results of which could be used as markers to identify distinct visual strategies associated with levels of performance.[3,9–11] For example, Ackerman *et al.* observed that experts developed efficient and refined strategies for accomplishing tasks, relative to novices.[12] Such attributes of experts could be explained using gaze-based metrics as employing minimal fixations[13,14] or limited scanning of the visual space[14,15] to achieve task objectives. Furthermore, Kok *et al.* identified systematic visual scanning as an attribute of expert performance, whereby different regions of interest (ROI) are scanned in a defined order.[8] In a recent review,[6] Brams *et al.* identified selective attention allocation, extended visual span and efficient visual search rate as attributes associated with expert performance. These features of eye gaze as a function of performance can be attributed to the ability of experts to understand informative (relevant) regions and ignore any irrelevant distractions. According to the information-reduction hypothesis,[16] as individuals gain expertise, they learn

to recognise relevant and irrelevant elements, enabling them to focus on processing relevant cues. The information-reduction hypothesis has been observed among art enthusiasts[17] and chess performers[18] where informative elements within a scene are static or less mobile. Indeed, studies exploring the relationship between visual strategies and proficiency levels often employed static visual scenes such as radiography images that include salient visual features.[8,17]

In the context of dynamic visual stimuli relatively less progress has been observed in the literature that explores differences in visual strategies associated with levels of proficiency.[19,20] In one study, Jarodzka *et al.* investigated eye movements based on fish locomotion, a visually complex and dynamic stimuli comprising of both the relevant and irrelevant information.[20] Salient features based on the locomotion description were used to distinguish the eye movements of novices from experts. Other perceptual-cognitive studies in dynamic ball sports have demonstrated that expert players react faster and more accurately compared to non-experts. According to Vansteenkiste *et al.* expert players were more successful in quickly analysing the trajectory of the ball in motion and predicted future location of the ball enabling superior performances.[21] Thus, it can be established that the presence of salient features in visual stimuli encourages the identification of visual strategies.

Task settings and objectives are critical when exploring the relationship between proficiency levels[6,20] and visual strategies because eye movements greatly rely on them.[22] The study presented in this paper attempts to analyse and evaluate the characteristics of eye movements in a dynamic setting where salient features are neither explicit nor obvious in the scene. For this purpose a simulated training environment with dynamic regions of interest (ROIs; i.e., moving unmanned aerial vehicles, UAVs)[23] is employed in the study. The primary objective was to refuel UAVs within specified time windows in order to prevent them from crashing. The task was designed using a repeated trial-based model to analyse the development of visual strategies. As participants progressed through trials they continuously gained knowledge about informative elements of the task. The skill acquisition of the participants through the trial progression was assessed using the task scores as an objective measure of performance.

To successfully achieve the task objective, participants were required to identify the fuel cycle of the UAVs which could be achieved by monitoring the fuel gauge and the fuel loss rate. The experimental task required efficient comprehension of the dynamics of the UAVs such as flying speed and the fuel consumption rates, whether they were similar or different, and the flight trajectories. We predicted that the underlying interpretations of these features would lead the participant to form a visual strategy to monitor and timely refuel the UAVs. The successful completion of the task was dependent on the participant's overall capability to comprehend the dynamics of the task structure, which demands a need for visual strategies due to the absence of the salient task features. Such an environment leads individuals to adopt differing visual strategies to achieve a common task goal.[24]

In the present study, we investigated the differing strategies adopted by participants while performing the UAV task.[25] Previous studies demonstrate the use of methods such as scan-path and heat-map estimation to understand the visual behaviours of individuals. However, the presence of both static and dynamic ROIs in the task means that scan-paths will not provide useful interpretation of eye gaze behaviours. As an alternative solution, epistemic network analysis (ENA)[26] was used to unfold the spatial organisation of fixations. The epistemic networks from trials were closely inspected to identify the strategies adopted by the participants. Cross recurrence quantification analysis (CRQA)[27–29] was employed to analyse temporal unfolding of gaze behaviours. An unsupervised clustering algorithm was employed on the CRQA measures to categorise gaze behaviours observed during trials into cases of strategy. This article proposes an eye-gaze metric based on the CRQA measures for quantifying the strength of visual strategies. The relationship between the strategies and the task performances was explored to assess the effects of gaze behaviour on task performance. Finally, we investigated which of these strategies were the most effective in achieving high performance outcomes within the defined task settings.

2 Methods

2.1 Participants

Twenty-five participants (age: mean = 32, standard deviation = 9, 23 males and 2 females) with normal or corrected-to-normal vision took part in the study. Each participant provided full written informed consent after reading the plain language statement and signing the consent forms approved by the Human Ethics Advisory Group (HEAG) of the Faculty of Science Engineering and Built Environment, Deakin University, Australia. All ethics approved by the HEAG complies with the human research ethics guidelines set by the Australian Code for the Responsible Conduct of Research and the National Statement on Ethical Conduct of Human Research.[30]

2.2 UAV Simulation Task

The objective of the task was to refuel a set of UAVs during an optimal refuel time window to prevent them from crashing. Fig. 1 shows an overhead "satellite" view of the visual display with three UAVs, each requiring refuelling at different time intervals. Participants were required to continuously monitor the fuel level of the UAVs and keep them flying for the maximum duration

possible. Points were awarded for correct actions and deducted for erroneous responses (details of awards and deductions are shown in Table. 1). A final score summarizing participants' performance was generated at the end of each trial.

Table 1. The UAV simulation task scoring scheme (for each instance)

Response	Awards
Correct-Refuel	1500
Correct-Hover-Over	25
Response	**Deductions**
Refuel-miss	-1250
False-refuel	-625
False-Hover-Over	-25

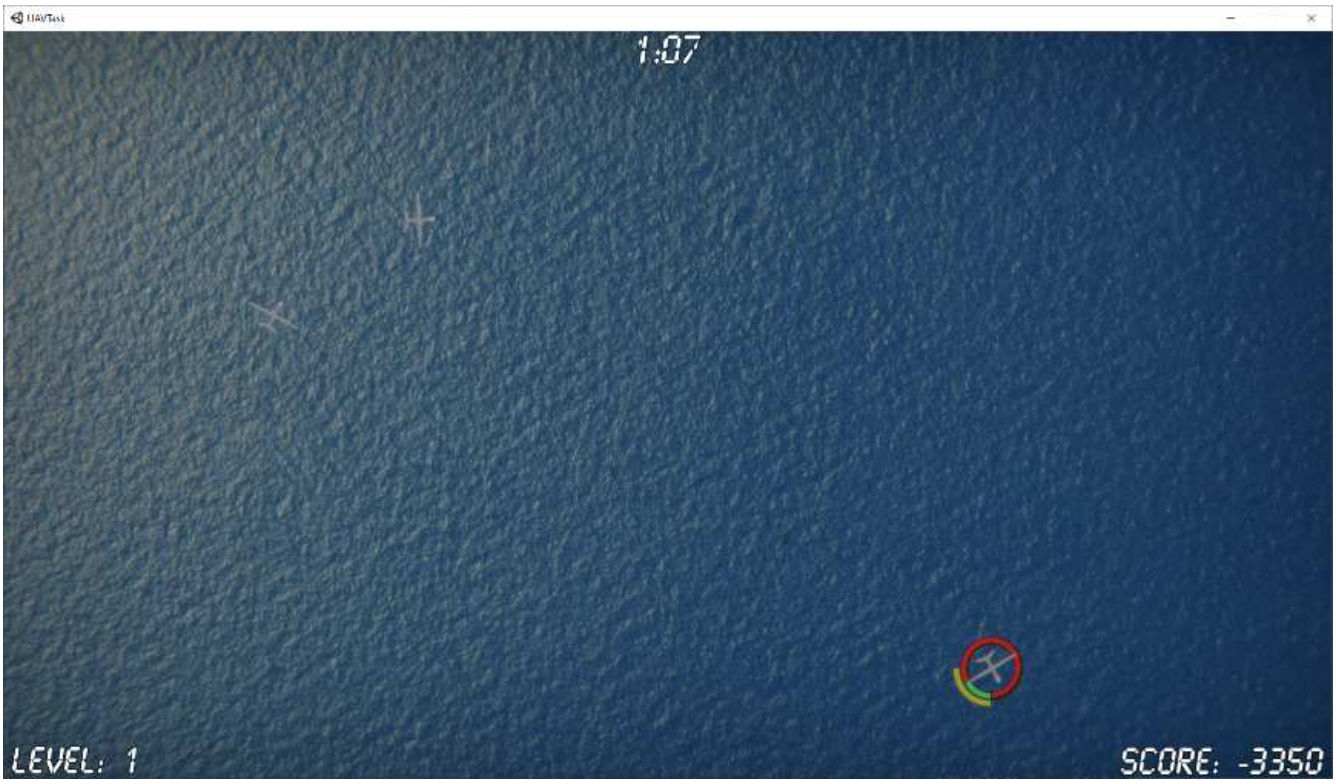

Figure 1. The UAV Simulation task displaying three UAVs adopted in the training session. When pacing a mouse curser over a UAV, its fuel level was displayed, showing how much fuel remaining (green colour), how much fuel has been consumed (red colour) and the optimal refuelling window (yellow colour).

Each UAV had a maximum fuel level of 100 arbitrary units (AU) with a fuel draining amount of 5 AU per second. In the absence of refuelling, they had 20-seconds of air time. At each successful refuel instance the UAV was set to its maximum fuel capacity. The optimal refuel time window was reached when the fuel level was $\leq$ 25 AU; that is, the UAV had an optimum refuel zone 15-20 seconds after it first appeared and after every correct refuel thereafter.

A pilot study was conducted to determine the appropriate level of task difficulty by varying the number of UAVs in trials between 3, 5 and 7. The result of the study showed that 5 UAVs per trial facilitated a moderate level of difficulty. Accordingly, in this experimental task, the number of UAVs was set to 5 per trial. The position of UAVs in (x, y) screen coordinates were recorded at 10 Hz sampling frequency. In addition, events corresponding to hover-over, refuelling, object-spawn, and object-destroy were also recorded.

2.3 Apparatus

A Gazepoint GP3 eye tracker[31] with 60 Hz sampling rate was used to capture the eye gaze data of the participants. The eye tracker was positioned under a 24-inch LCD monitor. Additionally, heart rate (HR) and galvanic skin response (GSR) were recorded using finger-based sensors.[25]

2.4 Experimental Procedure

Before starting the experiment, participants viewed an instructional video and participated in a two-minute training session comprising of three UAVs (see Fig. 1) to familiarise themselves with the task environment and objectives. Participants could clarify any questions they had regarding the task.

Upon completion of the training, the eye tracker was calibrated to each participant using the built-in 9-point calibration procedure. Participants then completed 10 repetitions of the experimental trials with each trial lasting two minutes. Participants were provided a break after five trials. Also, participants were informed they could take a break between trials if required. At the end of each trial, a feedback screen displaying their task performance such as total score, number of correct-refuels and number of refuel-misses was presented. After the completion of all trials, the participants completed a demographic questionnaire, which included questions such as age, gender, use of eye glasses, experience with eye-tracking and number of hours they used computers per day.

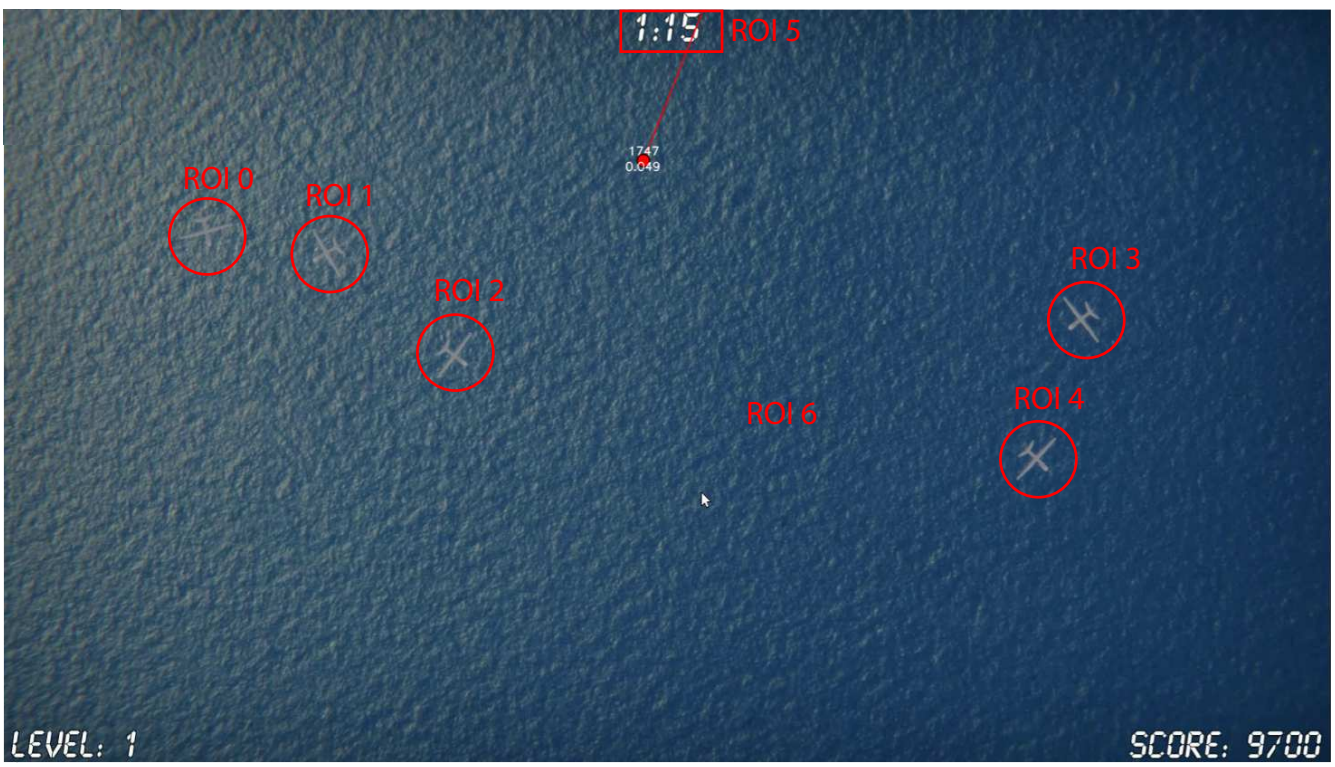

Figure 2. ROIs for a given time point $t(= 1:15\ minutes)$ in the simulation

2.5 Raw Gaze Data Processing

Raw gaze data was initially cleaned by eliminating samples where the eye-tracking system failed to track the pupils and iris of the participant. Gaze coordinates that exceeded 100 pixels outside of the calibrated region were also eliminated. The study employed cubic spline interpolation to replace the eliminated samples.

After cleaning the recorded gaze data, categorical gaze time-series were constructed from the gaze recorded in (x, y) screen coordinates. The regions of interest (ROIs) (Fig. 2) that we investigated with respect to participants eye-gaze included UAVs with random trajectories, the background display and the stationary clock. A dynamic ROI-based gaze time-series (Fig. 3 (a)) was constructed, as described below.

Let $ROI_i(x_t, y_t)$, $i = 0, 1, \ldots, 4$ be the dynamic ROIs corresponding to the UAVs at a given time t, centred at the position with the screen coordinates (x_t, y_t) and ROI_5 be the clock. At a given time point t the gaze fixation point $G(gx_t, gy_t)$, with the

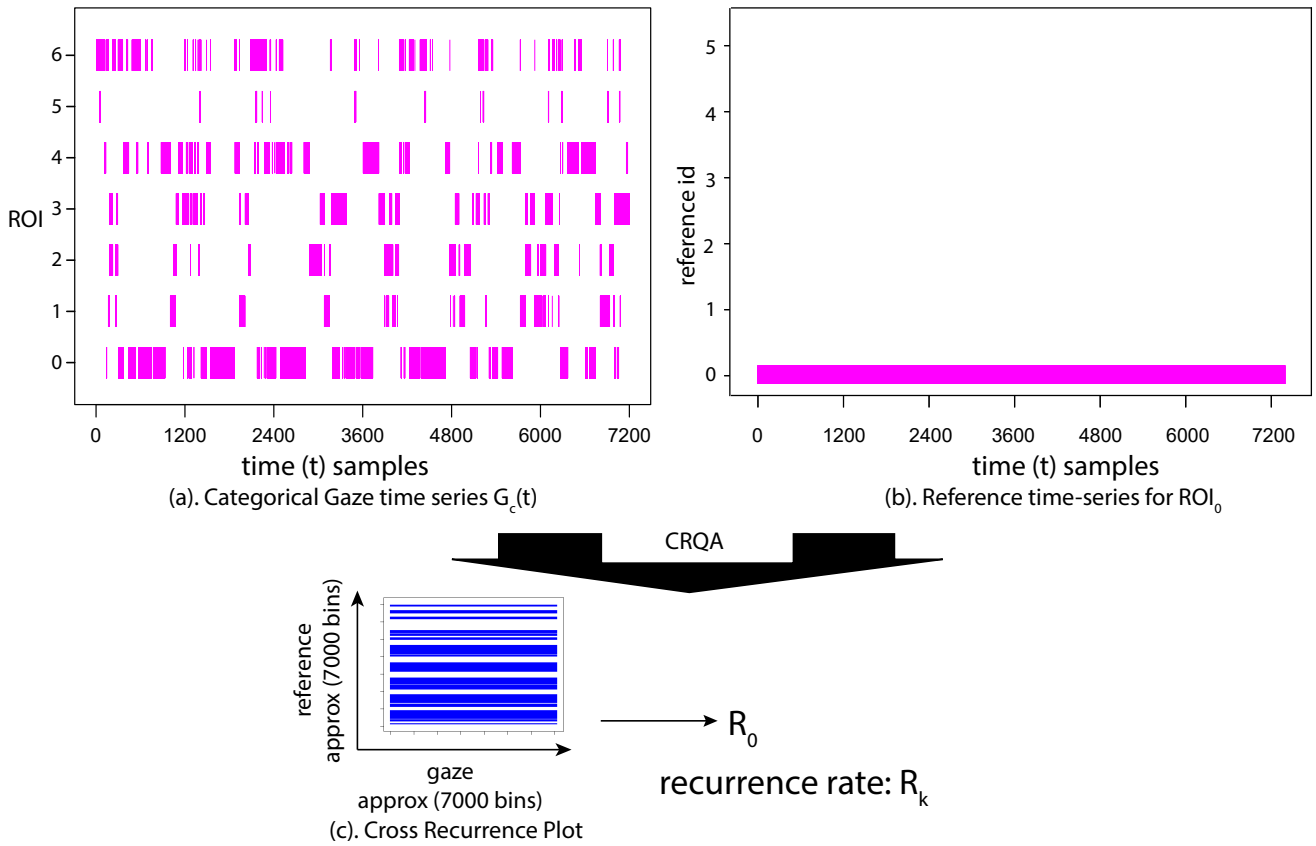

Figure 3. Example of evaluating recurrence rates R_k from categorical gaze time-series $G_c(t)$ (a) Example of $G_c(t)$ derived from eq. 3 for a participant's trial. (b) Reference time-series generated in relation with ROI_0. (c) The corresponding Cross Recurrence Plot (CRP) generated from CRQA.

screen coordinates (gx_t, gy_t) was assigned an ROI based on the Euclidean distance, d_i defined as,

$$d_i = ||ROI_i(x_t, y_t) - G(gx_t, gy_t)|| \text{ for } i = 0, 1, ..., 5 \tag{1}$$

Then, the ROI closest to the gaze point G was selected using,

$$\begin{aligned} i_{min} &= find(d_i == min(d_i)) \text{ for } i = 0, 1, \ldots, 5 \\ d_{min} &= d_{i_{min}} \end{aligned} \tag{2}$$

Based on a threshold ε the categorical gaze time-series was defined as,

$$G_c(t) = \begin{cases} i_{min}, & \text{if } d_{min} \leq \varepsilon \\ 6, & \text{otherwise} \end{cases} \tag{3}$$

where the threshold ε was set to 100 pixels in the current study.

The gaze time-series described above were used as input to the ENA and CRQA (details in section 2.6 and 2.7) methods to gain an understanding of underlying strategies adopted during tasks and to derive gaze-based measures. The analyses were carried out on the last 5 trials for all 25 participants eliminating any biases that could be accumulated due to any potential learning effects.

2.6 Epistemic Network Analysis (ENA)

ENA is an ethnographic technique for the identification and quantification of the relationship between elements of coded data using dynamic network representations.[26] ENA is widely used in the domain of learning analytics to investigate learning in

students who are actively involved in complex problem-solving.[26,32–34] The networks were used to identify and visualise interactions between cognitive elements associated with learning. The interactions represented strength of connection between the elements. The versatility of ENA enables the concept to be applied to analyse complex systems where a set of elements interact dynamically. In the context of eye-gaze analysis, ENA has been previously used to analyse spatio-temporal gaze coupling between dyads.[35] This study demonstrates the capability of ENA networks to successfully capture differences in eye movement coordination during different stages of the joint activity, including that of interaction sequences that resulted in breakdown and repair.

In the current study, patterns originating from ENA networks were used to guide meaningful representation of gaze behaviours. The networks represented gaze fixation patterns with respect to ROIs (elements) of the simulation environment. The networks were visually inspected to observe strategies developed by the participants. Interconnectivity between participants' ROI preferences were analysed to investigate the impact of different strategies on performance outcomes. The analysis was performed using the rENA package, which is an R-version of ENA.[26] Fig. 4 shows an example of the structure of the data used in the current analysis. The task activity code of each participant-trial consisted of a collection of epistemic frames, coded with ROIs from the simulation environment. An epistemic frame in this analysis, therefore, consisted of 7 frame elements (R0, R1,..., R6). The background was included in the analyses to understand the frequency of gaze transition in order to identify if there was a strategic scan pattern or whether the search pattern was random.

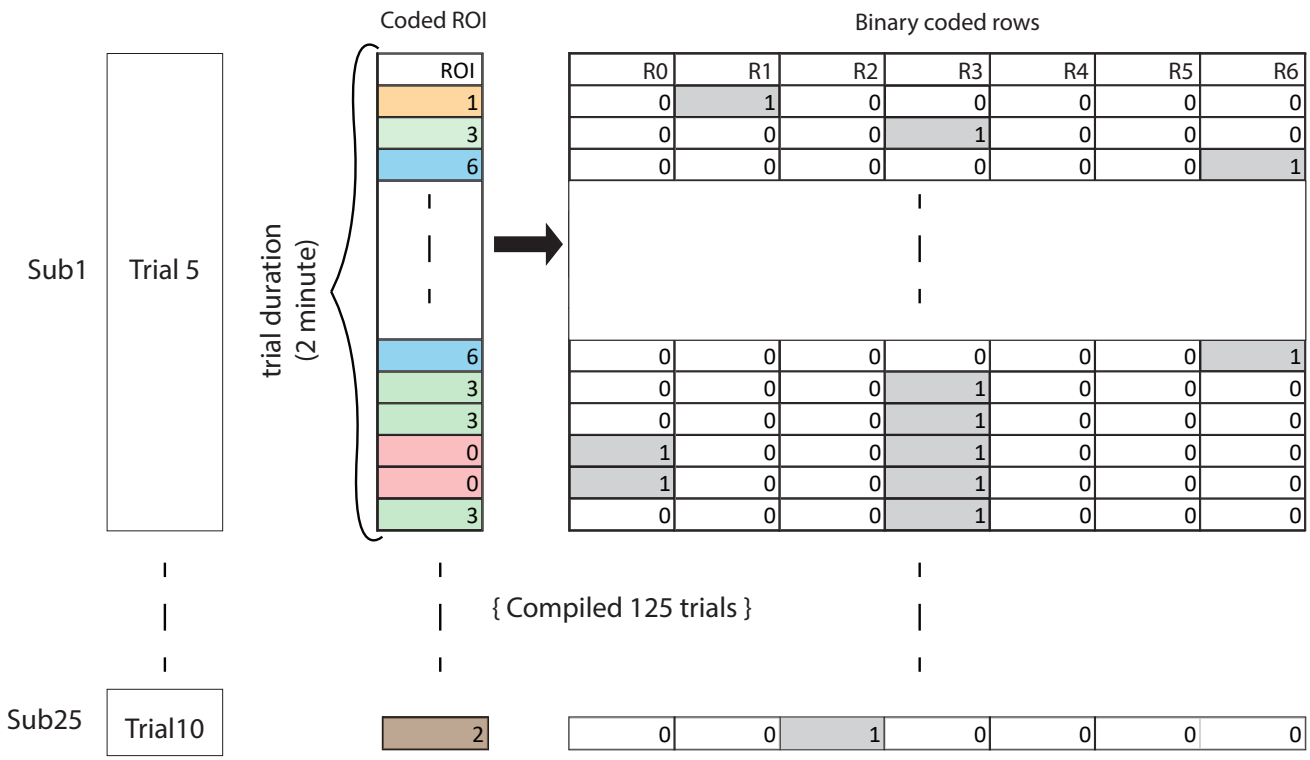

Figure 4. Data structure used for the ENA analysis. Coded ROI represents the concatenated gaze of all participants (1-25) and last five trials (5-10). The binary coded rows represent epistemic frame where references R0-R4 corresponds to UAVs, R5 is the clock reference, and R6 is reference for the background or elsewhere.

2.7 Quantification Analysis

CRQA is a nonlinear quantification tool that provides measures that can be used to analyse the temporal unfolding between two dynamically varying events.[36,37] CRQA was developed as an analytical tool for the purpose of studying human behaviour characteristics varying over time. Researchers in cognitive science have employed CRQA to quantify co-visitations when two people are involved in an interaction[38] such as the eye gaze of a dyad or postural sway between two individuals engaged in a conversation. Indeed, it has been previously used for studying and the quantification of gaze synchronisation during joint activities,[39–41] analysing coordination between pilots who are engaged in a team task,[42] and understanding performance of teams.[43] CRQA combines the concept of recurrence quantification analysis (RQA) to derive measures from CRP profiles such

as density of points, diagonal lines and vertical lines, which quantify co-visitations. Detailed information about CRQA may be found in the document by Coco and Dale.[38]

In the present study, CRQA was employed to understand the distribution of gaze on dynamic ROIs in an attempt to quantify the strategy underlying gaze behaviour. Thus, the analysis was used to understand the amount of gaze recurrence on each ROI during the 2-minute period of the trial. A CRQA measure, *recurrence rate* was used to distinguish performances on the basis of non-linear co-visitation between eye-gaze and ROIs. The pipeline of processing gaze data using CRQA is depicted in Fig. 3. Fig. 3(a) shows an example of categorical gaze time-series obtained in our study, plotting gaze fixations ROI_i at a discrete time (t). For example, $G_c(1210) = 5$ indicates that at a time sample $t = 1210$, gaze of the participant was fixated on ROI_5.

For each trial, the recurrence rate of gaze with each ROI was estimated as,

$$R_k(\tau) = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{R}_{i,j}(\tau) \tag{4}$$

where k represents ROIs ($\{0,1,...,6\}$ corresponding to UAVs, 5 represents the clock), N is number of samples and $\mathbb{R}(\tau)$ is the cross recurrence matrix with a radius of neighbourhood τ (assumed 0.001 in the current computation of R_k).

As an example, Fig. 3 shows calculation of recurrence rate R_0 (with relation to ROI_0) from $G_c(t)$. CRQA computes cross recurrence matrix $\mathbb{R}(\tau)$ as shown in Fig. 3(c) which is an asymmetrical representation of gaze recurrence between ROI_0 and its continuous time reference. The $\mathbb{R}(\tau)$ is further used to compute R_0 from eq. 4. The process is repeated for every ROI, i.e. for $k = \{0,1,...,6\}$ representing the gaze recurrence behaviour $\vec{R}$ of a participant's trial with a vector length of 7 as shown below

$$\vec{R} = R_0, R_1, R_2, R_3, R_4, R_5, R_6 \tag{5}$$

2.8 Visual Strategies

The recurrence behaviours observed in participant-trials were split into strategy cases to analyse the relationship between strategies and performance. An unsupervised clustering algorithm, Chameleon clustering, was employed to rationally split participant-trials into different cases.[44,45] Chameleon clustering is a hierarchical data splitting process that uses the k-nearest neighbours graph to identify clusters. Depending on the relative closeness between the clusters, the graphs are merged.

Accordingly, $\vec{R}$ of all 125 trials were first pooled together such that the parameters R_0 to R_6 represented features for the clustering process. The number of $K-nearest\ neighbours$ was set to 7. The parameter α was set to 2 prioritising relative closeness as opposed to relative interaction. Finally, the clustering process was repeated for different values of partitions ranging between $\{2,3,\ldots\ 7\}$. The result of each process was validated using Calinski-Harabasz (CH)[46] and Davies-Bouldin (DB)[47] indices to identify an appropriate partition.[48] It can be observed from the Fig 5 that at cluster-3, CH resulted the highest index and DB resulted in the lowest index. Thus, strategy cases obtained from 3-clusters are used in the remainder of the analysis.

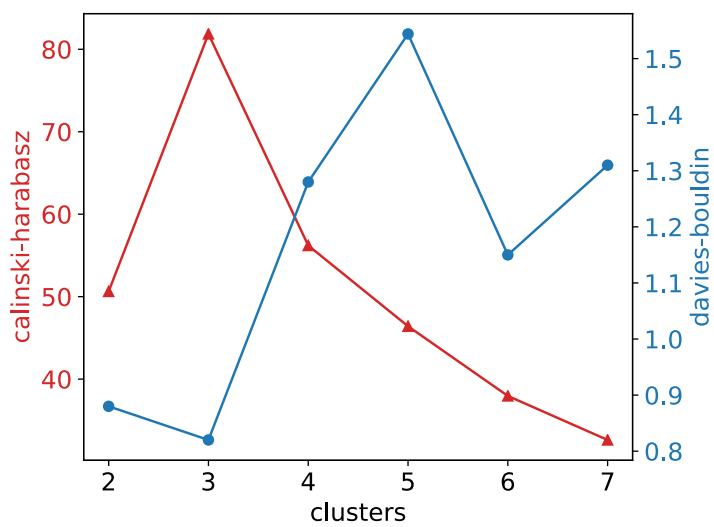

Figure 5. Result of CH and DB indices observed different clusters

2.9 Strength of Visual Strategy

An objective measure using $\vec{R}$ was used to represent the strength of visual strategy observed in a trial. It was observed that the time spent in the background only happened during the transition between the ROIs. Thus, gaze recurrence with the background does not directly impact a strategy. Hence, the gaze spent in the transition is not considered in the following analysis. The pool of $\vec{R}$ from 125-trials was subjected to statistical summary. On closer inspection of the relationship between summary statistics and performance score the following objective measure was derived. Accordingly, maximum recurrence rate was calculated as

$$\mathrm{R}_{max} = max(\mathrm{R}_0, \mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3, \mathrm{R}_4, \mathrm{R}_5), \tag{6}$$

and the median of recurrence rates was calculated as

$$\mathrm{R}_{median} = median(\mathrm{R}_0, \mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3, \mathrm{R}_4, \mathrm{R}_5), \tag{7}$$

and finally, the strength of a visual strategy R_{dev} as

$$R_{dev} = R_{max} - R_{median}. \tag{8}$$

Where, R_{dev} was a measure representing how much does R_{max} deviated from R_{median}.

3 Results and Discussion

The primary objective of the study was to critically investigate the relationship between visual strategies and task performance in a dynamic task setting. To conduct the investigations, gaze data from 25 participants, each performing 10 trials of a simulation task were recorded via desktop mounted eye-trackers.

The participants experience of the UAV simulation task was assessed based on the answers provided in the feedback questionnaire. On average, the participants scored 3.6 ± 1.5 (out of 7) for the challenging level of the task, indicating that the task difficulty was at a medium level. Further, the participants scored 6.46 ± 1 (out of 7) when answering the question 'I understood what was expected from me', which indicates that the participants had a good understanding about the task they had to perform during the simulation.

3.1 Performance-based Trial Grouping

The overall performance of a trial was reflected in feedback score. A probability plot (Fig. 6(A)) was constructed using the scores to distinguish between performances for evaluation in later stages of the analysis. As shown in Fig. 6, four different groups were identified, *outstanding, good, average* and *poor* based on the theoretical quantiles of probability plot. For each category, score thresholds, correct-refuels and refuel-misses, are summarised in Table. 2.

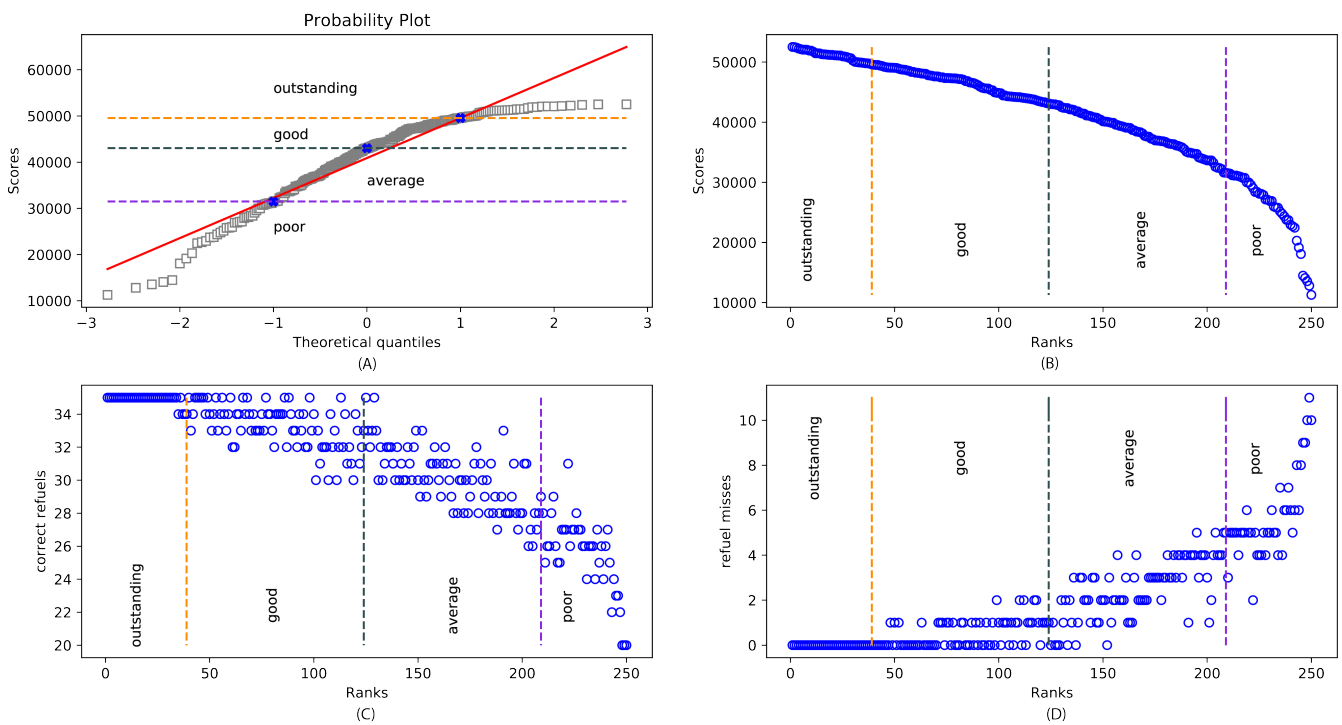

Figure 6. (A) Theoretical quantiles of probability plots used for distinguishing between performances. (B) Ranking of all participant-trials scores (overall performance). (C) Counts of correct-refuels (rewards) from corresponding to scores and rank. (D) Counts of refuel-misses (penalties) corresponding to scores and rank. The coloured lines orange, green and purple indicate thresholds, derived from the probability plots in shown in (A), used to establish performance categories

As can be observed from Fig. 6, participants with high scores were able to accurately refuel the UAVs with relatively smaller number of misses. Considering that first few trials could potentially be influenced by differences in learning rates

of participants,[49,50] to account for differences in visual strategies, the remainder of the analyses is only performed on the recordings from the later trials (trials 5 to 10).

Table 2. Score threshold and basis used for identifying performance categories

	Total	Outstanding	Good	Average	Poor
refuel-misses=1	45	0	32	13	0
refuel-misses=2	27	0	5	21	1
refuel-misses=3	23	0	0	23	0
refuel-misses≥4	61	0	0	22	39
refuel-misses=0 & correct-refuel<35	40	4	31	5	0
refuel-misses=0 & correct-refuel=35	54	36	17	1	0
Trials	250	40	85	85	40
Score threshold		52525 - 49550	49449 - 43012	43011 - 31468	31467 - 11275

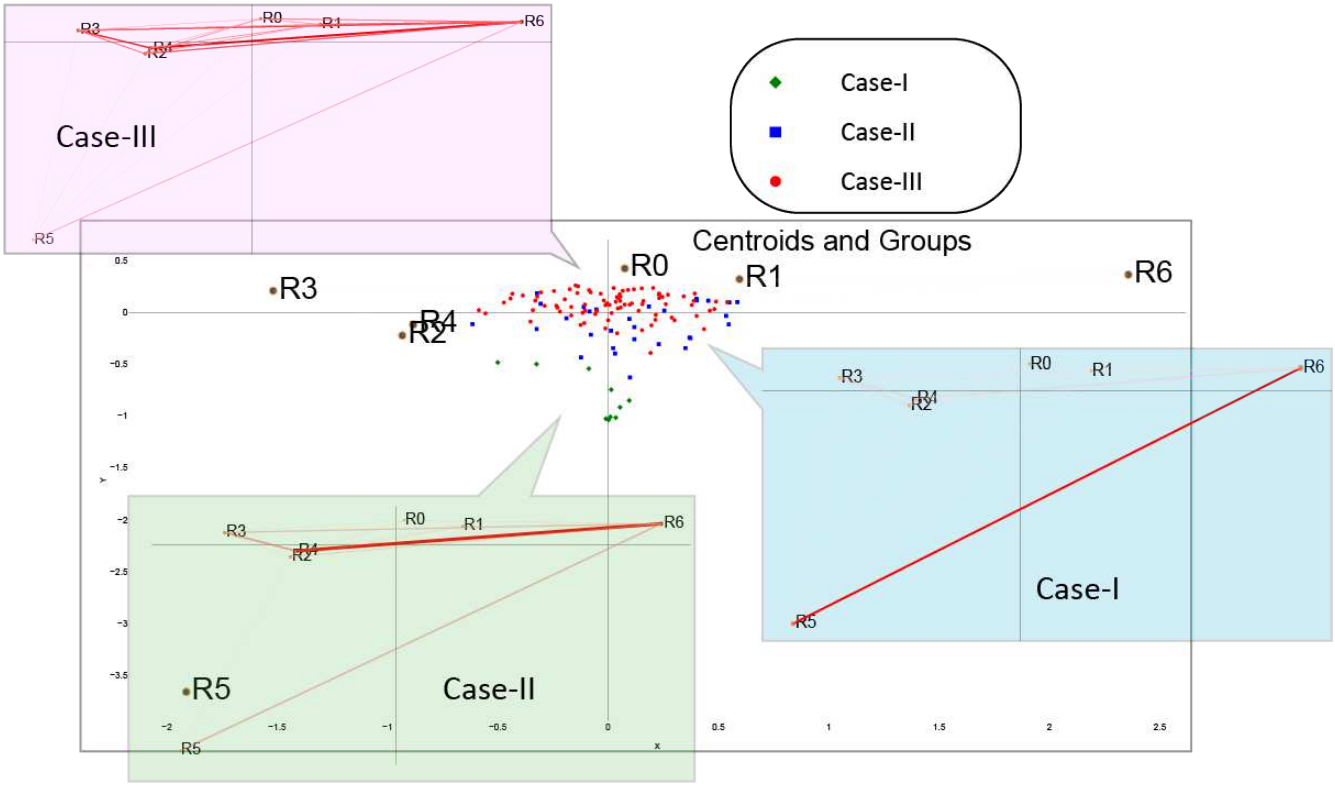

Figure 7. The centroids of each participant-trial combination on a single ENA space. Average networks corresponding to strategy cases are shown around the central plot.

3.2 Strategy Identification

ENA was used to analyse strategies adopted by participants during the simulation task. The centre plot in Fig. 7 shows the positioning of network centroids from 125 participant-trials on the ENA space. The nodes in the ENA space represent reference ROI, and the connection between nodes indicate the frequency of interaction. As eye-movement behaviour changes depending on a visual strategy, the connections between the nodes vary, positioning the centroids in different locations of the ENA-space.

Chameleon clustering was employed to differentiate the participant-trials into strategy cases. Following validation using internal indices (Fig. 5) the trials were split into three cases as shown in Fig. 7. Figure 7 also shows an average network representing each strategy group which provides insights of how interconnections between ROIs and how they can be used to distinguish strategy. Based on clustering and ENA networks, the three cases were identified as follows,

- Case-I: the individual ENA network shows strong connections (characterised by thick and dark connecting lines) between the clock (R5) and the background (R6), which is a result of longer gaze fixations on the clock relative to that of the rest of the interconnecting lines. Participants using this type of strategy likely relied on the fuel cycle times of the UAVs and therefore were reliant on the clock to guide their refuelling instincts. Participants that adopted this strategy appear to only momentarily leave clock position during refuel instances of the UAVs, meaning, they have spent considerable time in the background when switching their gaze between UAVs.

- Case-II: a strong connection between background (R6) and one of the UAVs (R4), can be observed in the ENA network. This means that the participant spent more time fixating on a single UAV (R4), relative to any other UAVs or ROIs. The participant may have been monitoring a particular UAV's fuel level, and at the refuelling instance, the fixated UAV was refuelled followed by rest of the UAVs and returned to their base position for the next refuelling instance.

- Case-III: the networks belonging to this category did not reveal any significant connections between the ROIs, compared to the previous cases (I and II). Some trials, which demonstrated randomness, had a very distributed gaze on UAVs and did not include the clock in their strategy at all. It appears that for this case, participants randomly kept track of the UAV positions. This might have been due to participants failing to learn the dynamics of the UAV simulation.

The results of this analysis provide an overall picture of how the ENA networks have efficiently extracted the unfolding of gaze patterns through shifts in interconnection frequency. The networks enabled comprehensive and closer inspection of the strategies adopted in each participant-trial. ENA was able to rationally (with respect to the task dynamics) characterise and distinguish the visual strategies adopted by the participants. This confirms our first prediction that, "visual strategies adopted by individuals" can be identified and categorised using gaze dynamics acquired via eye-tracking.

3.3 What is the Best Strategy?

A one-way ANOVA was performed to observe the relationship between performance and strategy cases and found that there were significant differences in task score as a function of the strategy cases (see Fig. 8). Tukey post-hoc tests with Bonferroni corrections revealed that the task scores in the strategy cases I and II were significantly higher than that of case-III ($F(2,125) = 19.76$, $p < .001$). There was no statistically significant difference between strategy cases I and II. The test revealed that the performance score was statistically significantly higher in case I (50091 ± 2749 , $p<.01$) and case II (48547 ± 3023, $p<.001$) compared to case III (40166 ± 8019). There was no statistically significant difference in task scores between strategy cases I and II. Thus, cases I and II were more effective in producing a higher score, and hence produced better task performance in the experiment.

From the above discussion and from Fig. 8(a) it can be inferred that an intended strategic approach guided by a reference ROI (cases I and II) contributed to higher performance scores. Thus, it can be confirmed that the development of strategies was essential to achieve an outstanding or good performance. Fig. 8(b) presents a more detailed break-down of scores as a function of task performance and the strategy used. The result can be summarised using performance categories *outstanding*, *good*, *average* and *poor* as,

- Participants that used the case-I strategy demonstrated relatively better performance than those that adopted alternative strategies. From Fig. 8(b) it can be observed that a large proportion of participant-trials employing this strategy demonstrated outstanding performance

- Of those participants that used case-II strategy, a large proportion demonstrated good performance

- Those participants that did not develop strategy indicated by the case-III were the only group to score poorly, with only a small proportion of this group producing outstanding performance.

- 70% of trials with case-I strategy produced *outstanding* scores, 20% of trials produced *good* performance, and 10% producing average score were considered outliers(Fig. 8(a)). Thus, the strategy where participants used the *clock* as reference were highly likely to produce *outstanding* scores and least likely to produce *poor* scores.

- 40% trials with case-II strategy produced *outstanding* scores and 58.06% produced *good* scores. Participant-trials employing the strategy demonstrated relatively less chances of producing *outstanding* scores compared to the proportion resulting in *good* scores. Furthermore, it is clear from the results that the case-II strategy has less chances of producing *poor* scores.

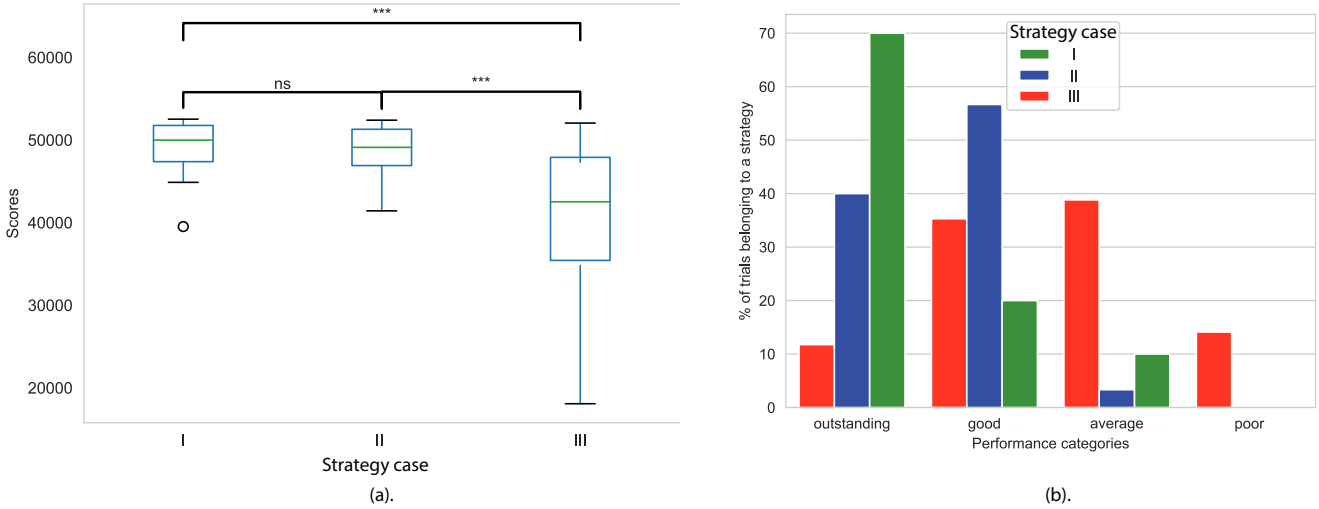

Figure 8. (a) Result of one-way ANOVA demonstrating the interrelationship of strategies cases as a function of performance scores. The significance test from Tukey post-hoc with Bonferroni correction can be interpreted using "***" which indicates p<.001 and "ns" indicates *no-significance*. (b) A detailed summary showing % of strategy cases contributing to a performance category

- Finally, having no strategy indicated in case-III meant that there are only limited chances of producing *outstanding* performance and is highly likely to produce *average* and *poor* scores.

Thus, the analyses establish that case-I guided by the *clock* is the most effective of all the strategies identified in our study and supports our assumptions that some strategies are better (more effective than others), and are linked to higher performance. Participants who followed the *clock* demonstrated high performance and a precise understanding of the task dynamics. Contrastingly, poor scores were the result of participants lacking an understanding of task structure or a motive to intuitively comprehend the task, as demonstrated in case-III strategies.

3.4 Relationship between Gaze Dynamics and Performance

Pearsons correlations were used to investigate the relationship between the gaze-behaviour and task performance (Fig. 9(a)). The deviation of maximum recurrence rate, R_{dev} demonstrated a significant positive correlation to the task performance scores (N=125, r = 0.46, p<.001). Thus, a participant-trial with relatively higher R_{max} on a certain ROI is highly likely to produce better task performance, demonstrating the presence of a potential strategy (belonging to either of strategy case I or II). On the contrary, the case-III trials would be characterised with a relatively lower R_{max} and hence a smaller deviation R_{dev}.

A one-way ANOVA found that R_{dev} varied significantly as a function of performance group F(3, 125) 11.58, $p < .001$. As can be seen in (Fig. 9), there was a linear trend in the data such that R_{dev} increased as a function of performance. Tukey post-hoc analyses with Bonferroni corrections found significant differences in R_{dev} between each of the performance groups, with the exception that there was no significant difference in R_{dev} between the average and poor performance groups.

3.5 General Discussion

This research investigated the influence of gaze-based attention strategies on task performance in nonlinear and dynamic task settings. The absence of salient features in the adopted simulated monitoring task environment meant that participants differed in their approach towards development and adoption of strategies as they progressed through the trials. This resulted in a diverse visual strategy pool producing different performance outcomes.

The correlation analysis and significance test established that, in the presence of a strategy, the corresponding trial is characterised by stronger gaze on one of the ROIs (longer fixation of the relevant region of visual space) and weaker gaze on the remaining ROIs (relatively fewer fixations on irrelevant regions). This strategic distribution of gaze was significant among high performing participant-trials. In line with the *information reduction* hypothesis,[16] the gaze strategies in case I and II were developed by participants after comprehensively understanding the dynamics of the task and the task structure. The deviation measure R_{dev} derived from *recurrent rates* was a compound representation of attributes of expert performances; that is, visual distinction of relevant and irrelevant information. Further, the statistical tests revealed that the strategies identified via the ENA were efficiently well complemented by R_{dev}. The analysis demonstrated that selective attention was predominant in the current

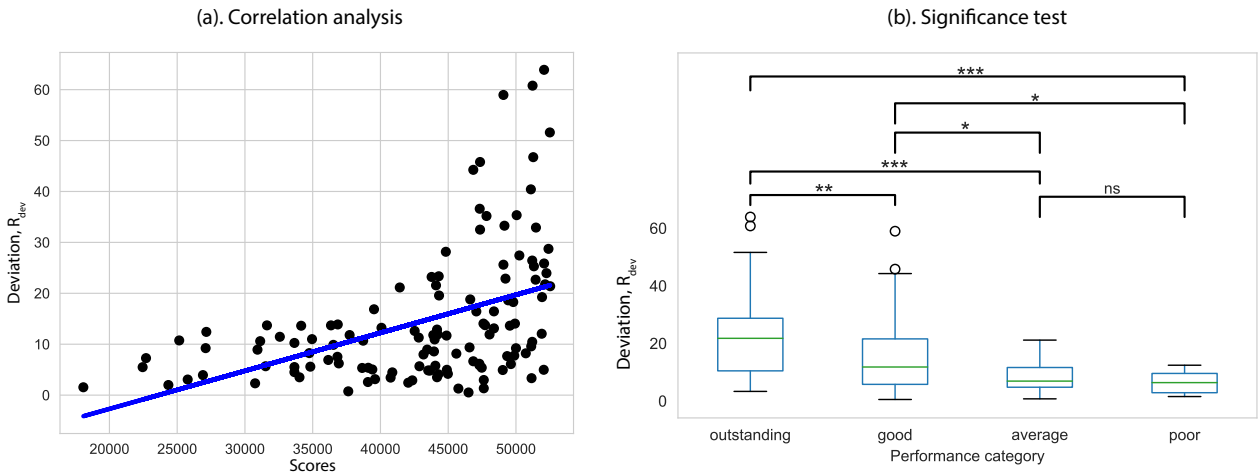

Figure 9. (a) Correlation analyses of R_{dev} (eq. 8) vs task performance scores. (b) one-way ANOVA analysis to observe significance of deviation measure (R_{dev}) among performance categories. Significance results p <.05 (*), p <.01 (**), p <.001 (***) and ns indicates no significance

task setting rather than systematic scanning. Thus, R_{dev} could be associated with selective attention, which formed the basis for many *outstanding* and *good* performances.

Overall, it can be established that visual strategies improve the chances of achieving better performance. The strategic behaviour was an emerging state due to participants gaining knowledge of the task structure and informative cues in the environment. The underlying idea of the observed strategies was to efficiently keep track of more than one entity at the same time, accumulating minimal errors. Despite the presence of diverse strategies to achieve *outstanding* scores, however, these scores were primarily motivated by selective attention allocation. The attention behaviours such as in case-III associated with *average* and *poor* performance can be more susceptible to errors. This is because case-III strategies could not be related with any of the performance-based gaze behaviour attributes such as selective attention, extended gaze or an efficient visual search rate.[6]

4 Conclusions

In the paradigm of task performance evaluation, this study presents methods for visualisation and quantification of gaze-behaviour. These methods efficiently analyse gaze behaviours observed in a dynamic task setting where salient features were less prominent. The ENA offered a robust visualisation technique for gaining better insights into visual strategies underlying a trial. Inspired by the theory of selective attention, the proposed quantification manifesting from an empirical approach can offer a better alternative to subjective measures. The effectiveness of ENA based dynamic representation to distinguish visual strategies and their relationship with CRQA measures will support an empirical approach that potentially can move towards more natural settings. Ultimately, the combination of ENA and CRQA provides an efficient visualisation, analysis and quantification platform for environments where informative elements are both dynamic and static in nature.

Furthermore, the much debated relationship between attributes of eye-movement patterns and performance is effectively addressed in this article, and this will aid studies investigating areas that need more attention.[6] Future research could focus on using decision tree algorithms to establish efficient gaze attributes that could prove valuable in dynamic environments. This framework provides a solid foundation for analysing mental processing attributes originating from visual perception in a dynamic environment such as visual strategy development and learning.

Acknowledgement

This research was funded by the Trusted Autonomous Systems - Strategic Research Initiative, Defence Science Technology Group, Australia.

The authors would also like to thank Professor Andrew Heathcote and Mr Matthew Gretton for sharing the source-code of the UAV simulation task. The UAV simulation task was extended and adapted, internally, at IISRI for the purpose of this study.

References

1. Benner, P. From novice to expert. *American Journal of nursing* **82**, 402–407 (1982).
2. Endsley, M. R., Garland, D. J. *et al.* Theoretical underpinnings of situation awareness: A critical review. *Situation awareness analysis and measurement* **1**, 24 (2000).
3. van Meeuwen, L. W. *et al.* Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction* **32**, 10–21 (2014).
4. Hauland, G. Measuring individual and team situation awareness during planning tasks in training of en route air traffic control. *The International Journal of Aviation Psychology* **18**, 290–304 (2008).
5. Ratwani, R. M., McCurry, J. M. & Trafton, J. G. Single operator, multiple robots: an eye movement based theoretic model of operator situation awareness. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, 235–242 (IEEE Press, 2010).
6. Brams, S. *et al.* The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological bulletin* **145**, 980 (2019).
7. Tien, T. *et al.* Eye tracking for skills assessment and training: a systematic review. *journal of surgical research* **191**, 169–178 (2014).
8. Kok, E. M. *et al.* Systematic viewing in radiology: seeing more, missing less? *Advances in Health Sciences Education* **21**, 189–205 (2016).
9. Koh, R. Y., Park, T., Wickens, C. D., Ong, L. T. & Chia, S. N. Differences in attentional strategies by novice and experienced operating theatre scrub nurses. *Journal of Experimental Psychology: Applied* **17**, 233 (2011).
10. Williams, A. M., Ward, P., Smeeton, N. J. & Allen, D. Developing anticipation skills in tennis using on-court instruction: Perception versus perception and action. *Journal of Applied Sport Psychology* **16**, 350–360 (2004).
11. Bertram, R., Helle, L., Kaakinen, J. K. & Svedström, E. The effect of expertise on eye movement behaviour in medical image perception. *PloS one* **8**, e66169 (2013).
12. Ackerman, P. L. & Beier, M. E. 13 methods for studying the structure of expertise: Psychometric approaches. *The Cambridge handbook of expertise and expert performance* 213 (2018).
13. Krupinski, E. A. *et al.* Eye-movement study and human performance using telepathology virtual slides. implications for medical education and differences with experience. *Human pathology* **37**, 1543–1556 (2006).
14. Manning, D., Ethell, S., Donovan, T. & Crawford, T. How do radiologists do it? the influence of experience and training on searching for chest nodules. *Radiography* **12**, 134–142 (2006).
15. Krupinski, E. A. Visual scanning patterns of radiologists searching mammograms. *Academic radiology* **3**, 137–144 (1996).
16. Haider, H. & Frensch, P. A. Eye movement during skill acquisition: more evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **25**, 172 (1999).
17. Antes, J. R. & Kristjanson, A. F. Discriminating artists from nonartists by their eye-fixation patterns. *Perceptual and motor Skills* **73**, 893–894 (1991).
18. Charness, N., Reingold, E. M., Pomplun, M. & Stampe, D. M. The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & cognition* **29**, 1146–1152 (2001).
19. Ellis, S. R. & Stark, L. Statistical dependency in visual scanning. *Human factors* **28**, 421–438 (1986).
20. Jarodzka, H., Scheiter, K., Gerjets, P. & Van Gog, T. In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction* **20**, 146–154 (2010).
21. Vansteenkiste, P., Vaeyens, R., Zeuwts, L., Philippaerts, R. & Lenoir, M. Cue usage in volleyball: A time course comparison of elite, intermediate and novice female players. *Biology of Sport* **31**, 295 (2014).
22. Yarbus, A. L. Eye movements during perception of complex objects. In *Eye movements and vision*, 171–211 (Springer, 1967).
23. Boehm, U. *et al.* Real-time prediction of fluctuations in cognitive workload (2019).
24. Medin, D. L. *et al.* Folkbiology of freshwater fish. *Cognition* **99**, 237–273 (2006).
25. Hettiarachchi, I. T. *et al.* Performance quantification and heart rate analysis in a repeated-trial simulation-based training task. In *2021 IEEE International Systems Conference (SysCon)*, 1–7 (IEEE).

26. Shaffer, D. W., Collier, W. & Ruis, A. A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics* **3**, 9–45 (2016).

27. Marwan, N. & Kurths, J. Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters, Section A: General, Atomic and Solid State Physics* **302**, 299–307 (2002).

28. Demir, M., McNeese, N. J. & Cooke, N. J. Team synchrony in human-autonomy teaming. In *International Conference on Applied Human Factors and Ergonomics*, 303–312 (Springer, 2017).

29. Kazi, S. *et al.* Team physiological dynamics: A critical review. *Human factors* 0018720819874160 (2019).

30. Human ethics advisory group (2021). URL https://www.deakin.edu.au/research/support-for-researchers/research-integrity/human-ethics.

31. Gazepoint gp3 (2009). URL https://www.gazept.com/.

32. Shaffer, D. & Ruis, A. *Epistemic Network Analysis: A Worked Example of Theory-Based Learning Analytics*, 175–187 (2017).

33. Hatfield, D. The right kind of telling: An analysis of feedback and learning in a journalism epistemic game. *International Journal of Gaming and Computer-Mediated Simulations* **7**, 1–23 (2015).

34. Arastoopour, G., Shaffer, D., Swiecki, Z., Ruis, A. & Chesler, N. Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education* **32**, 1492–1501 (2016).

35. Andrist, S., Collier, W., Gleicher, M., Mutlu, B. & Shaffer, D. Look together: Analyzing gaze coordination with epistemic network analysis. *Frontiers in psychology* **6**, 1016 (2015).

36. Dale, R., Kirkham, N. & Richardson, D. The dynamics of reference and shared visual attention. *Frontiers in Psychology* **2** (2011).

37. Veerabhadrappa, R., Hettiarachchi, I. T. & Bhatti, A. Using recurrence quantification analysis to quantify the physiological synchrony in dyadic ecg data. In *2021 IEEE International Systems Conference (SysCon)*, 1–8 (IEEE).

38. Coco, M. I. & Dale, R. Cross-recurrence quantification analysis of categorical and continuous time series: an r package. *Frontiers in Psychology* **5**, 510 (2014). URL https://www.frontiersin.org/article/10.3389/fpsyg.2014.00510.

39. Richardson, D. & Dale, R. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science* **29**, 1045–1060 (2005).

40. Richardson, D., Dale, R. & Kirkham, N. The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue: Research article. *Psychological Science* **18**, 407–413 (2007).

41. Richardson, D., Dale, R. & Tomlinson, J. Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science* **33**, 1468–1482 (2009).

42. Gontar, P. & Mulligan, J. Cross recurrence analysis as a measure of pilots' coordination strategy. In *32nd Conference of the European Association for Aviation Psychology* (2016).

43. Tolston, M. *et al.* Evaluating the relationship between team performance and joint attention with longitudinal multivariate mixed models. In *Proceedings of the Human Factors and Ergonomics Society*, 739–743 (2016).

44. Veerabhadrappa, R., Ul Hassan, M., Zhang, J. & Bhatti, A. Compatibility evaluation of clustering algorithms for contemporary extracellular neural spike sorting. *Frontiers in Systems Neuroscience* **14**, 34 (2020). URL https://www.frontiersin.org/article/10.3389/fnsys.2020.00034.

45. Karypis, G., Han, E.-H. & Kumar, V. Chameleon: Hierarchical clustering using dynamic modeling. *Computer* **32**, 68–75 (1999).

46. Caliński, T. & Harabasz, J. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* **3**, 1–27 (1974).

47. Davies, D. L. & Bouldin, D. W. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence* 224–227 (1979).

48. Zhang, J. *et al.* A review on cluster estimation methods and their application to neural spike data. *Journal of neural engineering* **15**, 031003 (2018).

49. Brockmole, J. R., Hambrick, D. Z., Windisch, D. J. & Henderson, J. M. The role of meaning in contextual cueing: Evidence from chess expertise. *The Quarterly Journal of Experimental Psychology* **61**, 1886–1896 (2008).

50. Newell, A. & Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition* **1**, 1–55 (1981).