

1 **Gaze behaviours, situation awareness, and cognitive workload of air**
2 **traffic controllers in radar screen monitoring tasks with varying task**
3 **complexity**

4 Cho Yin Yiu ^a, Kam K.H. Ng ^{a*}, Qinbiao Li ^a and Xin Yuan ^a

5 *^aHuman Factors and Ergonomics Laboratory, Department of Aeronautical and*
6 *Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China*

7 Email Address: james-cho-yin.yiu@connect.polyu.hk (Cho Yin Yiu),
8 kam.kh.ng@polyu.edu.hk (Kam K.H. Ng), qinbiao.li@connect.polyu.hk (Qinbiao Li),
9 xin-olivia.yuan@connect.polyu.hk (Xin Yuan).

10 **Acknowledgment**

11 The research is supported by *Department of Aeronautical and Aviation Engineering,*
12 *The Hong Kong Polytechnic University, Hong Kong SAR.* Our gratitude is also extended
13 to the Research Committee of the *Department of Aeronautical and Aviation*
14 *Engineering, The Hong Kong Polytechnic University* for support of the project (RLPA,
15 CE1G, RH1W, RJX2) and the Hong Kong PhD Fellowship (Reference number: PF21-
16 62058). This study has been granted human ethics approval from the PolyU Institutional
17 Review Board of *The Hong Kong Polytechnic University* (IRB Reference Number:
18 HSEARS20210318002).

19

20 **Gaze behaviours, situation awareness, and cognitive workload of air**
21 **traffic controllers in radar screen monitoring tasks with varying task**
22 **complexity**

23 **Abstract**

24 **Purpose:** Air traffic controllers shall maintain a high SA and low cognitive
25 workload to ensure aviation safety. However, the increased task complexity may
26 influence air traffic controllers' SA and cognitive workload. Meanwhile, eye-
27 tracking provides insights into the gaze patterns that might signify SA. This paper
28 investigates the gaze behaviours, situational awareness (SA), and cognitive
29 workload of different radar screen monitoring tasks with varying task
30 complexity.

31 **Materials and methods:** Twenty-eight participants performed three radar screen
32 monitoring tasks, including callsign association, position identification, and
33 heading projection. The cognitive workload and SA are evaluated for each task
34 using NASA-TLX and SAGAT, respectively. The Gaussian Mixture Model was
35 used to cluster the SA into high/low. Eye-tracking reveals the gaze behaviours
36 that may contribute to the SA formation.

37 **Results and conclusions:** The cognitive workload and SA significantly differ
38 between different levels of task complexity. While task complexity has a
39 significant main effect on the fixations of human operators, it does not
40 significantly influence the pupil diameter. This study concludes that fixation-
41 related metrics may signify the changes in situational awareness under varying
42 task complexity, while the side effects of cognitive workload shall be mitigated.

43 **Keywords:** Eye-tracking; situational awareness; cognitive workload; air traffic
44 control; neuroergonomics

45

46 **1. Introduction**

47 As the regulators of the airspace, air traffic controllers (ATCOs) are responsible for
48 resolving potential conflicts and providing flight guidance and navigation information
49 to pilots by monitoring flight parameters to make decisions [1-3]. Past research
50 suggested that the primary elements that ATCOs should recognise are callsign, position,
51 and heading of the aircraft within their area of jurisdiction [4,5]. As the air traffic
52 environment is a real-time system, so they need to make an appropriate near-time
53 decision, which is a challenging, demanding, and stressful especially with a post-
54 pandemic surge in the number of flights [6-8]. Therefore, ATCOs shall monitor these
55 elements from the radar map environment to support their decisions and air traffic
56 control safety [9,10].

57

58 Indeed, numerous accidents in the past have demonstrated the importance of ATCOs
59 maintaining their situational awareness (SA) [11]. For instance, in Prinair Flight 277,
60 the ATCO perceived the aircraft to be positioned 10 miles further west than its actual
61 position, which indicates an incorrect SA. The ATCO gave an incorrect instruction
62 causing the aircraft crashed into a mountain with 19 casualties. Lufthansa flight 527 is
63 another fatal example of ATCO not being attentive towards the aircraft and failing to
64 realise the aircraft's critical distance from the terrain. The ATCO maintained his
65 attention on another aircraft while ignoring the Lufthansa flight. Hence, ATCOs losing
66 SA may provide pilots with wrong guidance and instructions, greatly impacting flight
67 safety [12,13]. Therefore, the cognitive workload and SA of ATCOs shall be regularly
68 monitored so corrective actions can be made whenever required [14].

69

70 Given the ability to reveal the allocation of attentional resources, eye-trackers have been
71 used to monitor and infer SA in various domains. As human operators need to gain SA
72 based on the primary elements, this research aims to explore how gaze behaviours,
73 cognitive workload, and SA differ in tasks with varying complexity. We specifically
74 focus on the investigation of non-ATCOs sample to investigate the tasks without the
75 influence of training and experiences. The use of naïve participants with basic ATC
76 knowledge provides better insight into workload and the real difficulty of the task. The
77 gaze behaviours shall give implications on how SA is formulated in different task
78 requirements and thus explore if eye-tracking can be used to monitor the SA of ATCOs
79 to secure flight safety actively. Specifically, the knowledge gained from this study may
80 have implications on ATCOs training based on the results from the naïve participants.
81 The rest of the paper is organised as follows. The next section reviews SA and related
82 eye-tracking studies. The methods section describes the experimental design, procedure,
83 and method of analysis. Then, results and discussion are presented. The conclusions,
84 limitations, and recommendations for future work are described in the last section.

85 **2. Related works**

86 ***2.1. Situational awareness, cognitive workload, and task complexity***

87 SA refers to having a good understanding of what is happening in the surrounding
88 environment [15]. [Endsley \[16\]](#) divided SA into three levels: SA Level 1 requires ‘the
89 perception of environmental elements and events in time and space’; Level 2 requires
90 ‘the comprehension of the current situation with an understanding of the significance of
91 the elements to the task goals’; and Level 3 requires ‘the project of their future status
92 with knowledge of the status’. Therefore, SA demands the ability to update and extract
93 meaningful information, the primary elements that help them to make appropriate

94 decisions. Among the three SA levels, SA Level 1 is the cornerstone for the two higher
95 SA levels and is the most essential for ATCOs. Indeed, task complexity may also
96 influence the achievable SA due to the difference in cognitive workload [17-19]. For
97 instance, the callsign of flights displayed on the radar screen are limited and can be
98 obtained from the flight progress strips. ATCOs with known flight positions only need
99 to associate the appropriate callsign with the position, which is less cognitively
100 demanding. Conversely, ATCOs may need to process the information gained during
101 monitoring to locate a given callsign and its heading because there are nearly infinite
102 possibilities of flight position and heading. They may find it more cognitively
103 demanding, while human performance may drop under a high cognitive workload.
104 Thus, we hypothesise the following:

105 **H1.** The cognitive workload significantly differs between the three radar screen
106 monitoring tasks at different levels of complexity.

107

108 Furthermore, traditional SA assessments are conducted in objective probes or subjective
109 self-rated questionnaires [20]. The Situational Awareness Global Assessment Technique
110 (SAGAT) adopts a freeze-probe technique during goal-directed tasks to query the
111 operator about task-relevant knowledge [21]. The percentage of correct answers
112 throughout repetitive task trials is used to determine the operator's SA. However, this
113 strategy could hardly facilitate real-life applications as a 'freeze' is impossible [20].

114 Assuming one knows where to obtain task-relevant information in the environment,
115 another approach called the Situation Present Assessment Method (SPAM) emphasises
116 the response latency [22]. However, these assessments are obtrusive, and one needs to
117 accomplish additional sub-tasks for measurement [23]. It hindered instantaneous SA
118 measurements as the sub-task may cause distraction, which shall be transformed into

119 non-distracting active monitoring. In this study, we employ SAGAT to validate the SA
120 measurement. The drop in human performance under a high cognitive workload may
121 hinder ATCOs from gaining SA from the environment. We thus hypothesise that:
122 **H₂**. The SAGAT scores significantly differ between the three radar screen monitoring
123 tasks at different levels of complexity.

124 *2.2. Eye-tracking and fixation-based metrics*

125 Eye-trackers, as an indicator for attentional resource allocation, have been used to infer
126 SA in different areas of application in performance monitoring [24]. [Zhang, Yang \[25\]](#)
127 suggested that spending a longer time at an AOI for information gathering would gain a
128 better understanding of that environment. [Nguyen, Lim \[20\]](#) also defined the SA
129 assessment via tracked eye movements as process indices. Several empirical studies
130 presented positive correlations between conscious eye-tracking metrics and SA [26-29].
131 [Paletta, Dini \[30\]](#) suggested that Spearman's rank correlation coefficient is a significant
132 factor ($\rho = 0.608$; $p < 0.05$) between the fixation time and SAGAT score.
133 [Hasanzadeh, Esmaeili \[26\]](#) found that higher fixation counts and longer dwell time
134 positively correlate with SA. [Moore and Gugerty \[27\]](#) suggested that participants with
135 higher SAGAT scores have a higher fixation count. The results of the above studies lay
136 a solid empirical ground for eye-tracking metrics, especially on fixation counts.

137
138 Furthermore, recent studies employ eye-trackers to study SA for the safety of
139 conditionally automated driving. [Liang, Yang \[15\]](#) investigated the effects of pre-
140 takeover visual engagement on SAGAT scores, pupil size, fixation durations, and
141 saccade counts. The results demonstrated that pre-takeover visual engagement
142 significantly influences the SAGAT score, pupil size, and total fixation duration.
143 Furthermore, [Zhou, Yang \[31\]](#) predicted SA during takeover transitions using the Light

144 Gradient Boosting Machine (LightGBM) with eye-tracking metrics input, which
145 evidenced that eye-tracking metrics may be a precursor of SA. In addition, [de Winter,](#)
146 [Eisma \[32\]](#) suggested that eye-tracking metrics are more suitable than SAGAT to
147 predict task performance. Eye-tracking is thus advocated as an indirect and objective
148 measure of SA. Given that the tasks requiring higher SA levels may demand cognitive
149 resources, the following are hypothesised:

150 **H₃**. The **(a)** total duration, **(b)** number of whole fixations, and **(c)** pupil diameter
151 significantly differ between the three radar screen monitoring tasks at different levels of
152 complexity.

153 **2.3. Research gap**

154 The current research challenge is that traditional SA assessment methods cannot satisfy
155 the need for real-time human performance monitoring. For example, SAGAT merely
156 measures the percentage of correctly answered probes. Its measurement needs an
157 aggregate of multiple probes, which cannot satisfy real-time assessment. Furthermore,
158 no concrete criteria define ‘high’ and ‘low’ SA, even based on traditional assessment
159 methods. Thus, we employ unsupervised learning to statistically distinguish between
160 high and low SA. On the other hand, numerous studies from other domains
161 demonstrated that eye-trackers can discriminate the allocation of visual attention
162 resources, which may reflect cognitive workload and SA changes in different task
163 complexity. Therefore, this research aims to reveal the gaze behaviours on how naïve
164 participants with basic ATC knowledge gain SA from the radar map environment.
165 Simultaneously, the NASA-TLX is utilised to provide additional subjective evidence to
166 understand the potential impacts of workload. Thus, the current research contributes to
167 developing a real-time indirect SA assessment technique based on eye-tracking.

168 3. Methods

169 3.1. Participants

170 During the experiment, 31 participants were recruited from the students and staff of the
171 Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic
172 University. A G-Power analysis with effect size 0.4 and $\alpha = 0.5$ suggested that a total
173 sample size of at least 18 participants is required. The study is also ethically approved
174 by the PolyU Institutional Review Board (Reference number: HSEARS20210318002).
175 Pre-experiment written informed consent was obtained. Participants were compensated
176 with monetary incentives. The experiment lasted for around 45 minutes. Due to a low
177 gaze sample accuracy (less than 60%), data from three participants were excluded from
178 the analysis. The participants were between 20 and 32 (M: 23.4; SD: 3.31). Normal
179 (42.9%) or corrected eyesight (57.1%) was achieved to mitigate the effects of vision
180 impairment. All participants can understand basic aviation terminologies and basic ATC
181 procedures. They were recruited from the Department of Aeronautical and Aviation
182 Engineering and took an introductory course on air traffic management. Hence, none of
183 the participants are ATCOs to prevent the influence of the experience and training.

184 3.2. Apparatus and measures

185 The study was conducted in the Human Factors and Ergonomics Laboratory,
186 Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic
187 University. The ATC simulation and the direct SA measurement were carried out in E-
188 Prime 3.0 on a 27-inch monitor (Participant's Monitor, resolution: 1920×1080 pixels,
189 refresh rate: 60 Hz, **Titles of Figures**

190 **Figure 1**). The eye-screen distance was approximately 83 cm. The room lighting was
191 the same for all participants. There was no sunlight as the room is enclosed without

192 windows. Only controlled lightings were available in the room with the same setting
193 across participants. The screen brightness was kept constant. The room dedicated to the
194 experiment was quiet, with all other potential stimuli removed [33]. The modified
195 NASA-Task Load Index (TLX) was adopted to assess the subjective workload of the
196 participants [34]. Another 24-inch monitor (Researcher's Monitor) was set up for real-
197 time monitoring. The screen-based eye-tracker Tobii Pro Fusion (Tobii, Sweden), with
198 a sampling frequency of 256 Hz, precision at 0.04°, and accuracy at 0.3°, was deployed
199 to record eye movements through Tobii Pro Lab.

200

201

[Figure 1 near here]

202 **3.3. Experimental design**

203 Based on the psychology experiment building language (PEBL) situational awareness
204 test [35], participants were required to complete three radar screen monitoring tasks:
205 callsign association, position identification, and heading projection. The flight scenarios
206 were modelled from real Hong Kong International Airport flight data. Each trial has a
207 five-second monitoring period (**Figure 2**) and a response period. In each trial, three
208 flights with different callsigns were presented on the screen and moved according to
209 their actual trajectories. The simulation freezes after the monitoring period. Participants
210 were single-blinded on the correctness of their answers.

211

212

[Figure 2 near here]

213 **3.3.1. Callsign association task**

214 The callsign association task requires the participants to correctly associate the callsign
215 of the captioned flight, demonstrating the ability to perceive the primary elements. The

216 position of one of the flights is shown in the response period (**Figure 3**), requiring the
217 participants to associate the correct callsign of the flight shown out of the four provided
218 choices. The correctness of the answer is recorded for SAGAT score calculation.
219 Answering correctly is considered as the participant achieving the required SA, and vice
220 versa. This task is considered the least complex, as ATCOs with known flight positions
221 only need to associate the appropriate callsign with the position shown.

222

223 [Figure 3 near here]

224 3.3.2. *Position identification task*

225 The flight position identification task requires the participants to comprehend the
226 information in the monitoring period to identify the position of a given flight. During
227 the response period, the participants are required to click on the last position of the
228 flight with the callsign displayed (**Figure 4**). The position error is recorded based on the
229 Euclidean distance between the correct and the selected position. This task is regarded
230 as a medium-complex task as ATCOs are only given a callsign but need to process the
231 information gained during monitoring to locate the flight position.

232

233 [Figure 4 near here]

234 3.3.3. *Heading projection task*

235 The heading projection task requires the participants to comprehend the information in
236 the monitoring period to project the heading of a given flight. During the response
237 period, they indicate the correct position and then click on a point about their perceived
238 aircraft heading (**Figure 5**). A green line connecting the selected position and the
239 selected point to indicate the heading. The position and heading errors, as defined by the

240 Euclidean distance and the angular difference between the correct and the selected
241 position, are recorded, respectively. This task is regarded as the most complex task as
242 ATCOs are only given a callsign but need to process the information gained during
243 monitoring to simultaneously locate the flight position and the heading.

244

245 [Figure 5 near here]

246 *3.4. Experimental procedure*

247 **Figure 6** outlines the experimental procedure.

248

249 [Figure 6 near here]

250

251 Upon arrival, participants read the information sheet and sign the consent form. Then,
252 the researchers briefed them about the experiment. The eye-tracker is calibrated using
253 the internal calibration algorithm of Tobii Pro Lab. Participants attempt six unrecorded
254 familiarisation trials per task before data collection. The actual experiment consists of
255 25 trials per task. Five seconds of monitoring time is deemed sufficient for ATCOs to
256 make near-time conflict resolution decisions. The tasks and each task trial are shuffled
257 and random across all participants. As the tasks are static, the ‘physical demand’ is less
258 likely to contribute to their cognitive workload, so it is removed from the NASA-TLX
259 questionnaire. After each test block, participants complete the modified NASA-TLX
260 questionnaire to assess their cognitive workload. Upon completion, participants receive
261 their monetary reward for their participation.

262 **3.5. Data analysis**

263 The monitoring period was selected as the time of interest (TOI). Then, fixation-based
264 metrics, including total duration, the number of whole fixations, and pupil diameter,
265 were extracted. The SAGAT scores are computed and clustered using the Gaussian
266 Mixture Model. The statistical analyses were conducted in IBM SPSS Statistics version
267 26.0. One-way repeated measures Analysis of Variance (ANOVA) was used to compare
268 the three tasks on different dependent variables mentioned. **Figure 7** indicates the data
269 analysis procedure.

270

271 [Figure 7 near here]

272 **4. Results**

273 **4.1. Cognitive workload**

274 First, we conduct a reliability analysis to ensure the items measure the same construct.
275 The Cronbach's Alpha (α) obtained for the callsign association task, position
276 identification task, and heading projection task are 0.756, 0.831, and 0.851,
277 respectively, which is higher than the acceptable threshold of 0.7 that [Hair \[36\]](#)
278 suggested. Then, the one-way repeated measures ANOVA is used to compare the
279 cognitive workload assessed in different tasks.

280

281 **Table 1**, **Table 2**, and **Table 3** show the descriptive statistics and the results of one-way
282 repeated measures ANOVA on NASA-TLX scores, respectively.

283

284 [Table 1 near here]

285 [Table 2 near here]

[Table 3 near here]

286

287

288 All domains (including unweighted average) of the NASA-TLX attained $p < 0.001$,
289 implying the cognitive workload significantly differs between three radar screen
290 monitoring tasks with varying task complexity. The pairwise comparison revealed that
291 the callsign association task incurs the least workload, followed by the position
292 identification and heading projection tasks. Thus, H_1 is confirmed.

293 4.2. SAGAT

294 4.2.1. Scoring

295 In the following, we present the calculation of the SAGAT score and the clustering into
296 high/low SA. In the callsign association task, Equation (1) and (2) are used to calculate
297 the SAGAT score. For each participant i , we divide the number of correct responses by
298 the total number of trials n . There are 649/700 trials (92.7%) that show a high SA
299 ($SA_{ij} = 1$).

300

$$SAGAT\ Score = \frac{\sum_{j=1}^n SA_{ij}}{n} \quad (1)$$

$$SA_{ij} = \begin{cases} 1, & \text{if the response from participant } i \text{ on } j^{th} \text{ trial is correct} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

301

302 On the other hand, the position identification and heading projection tasks are not
303 answered in the form of a binary value (correct/wrong) but in position error (in pixels)
304 and angular error (in $^\circ$). Therefore, we adopt the two-component GMM to model the
305 probability distribution of the error values. It clusters the samples into two components
306 – high SA and low SA without subjective judgment.

307

308 As GMM relies on the statistical distribution, outliers shall first be removed from the
309 clustering processes. Regarding the position error, we first compute the Euclidean
310 distance between the selected position and all three flights. The trials are labelled as low
311 SA and excluded from GMM if the computed distance is the lowest with a wrong flight.
312 Second, we compare the position error value among two tasks (Position identification:
313 227.8991 pixels; Heading projection: 515.6326 pixels) and select the lower value
314 (227.8991) as a criterion. The position errors greater than the criterion are also labelled
315 low SA and excluded from GMM. Furthermore, the angular errors are recorded
316 clockwise between $[0^\circ, 360^\circ]$. Thus, we subtract angular errors greater than 180° from
317 360° to ensure the range is limited to a semi-circle. We label trials with an absolute
318 angular error value greater than 90° as low SA and excluded from GMM. The
319 remaining trials are clustered using GMM on Python 3.7.9.

320 4.2.2. *Gaussian Mixture Model*

321 First, the position errors in the position identification task are clustered using GMM to
322 differentiate high and low SA. The probability density function (PDF) fitted by the
323 GMM is shown in **Figure 8**.

324

325 [Figure 8 near here]

326

327 From **Figure 8**, the intersection point e' between the PDF for 'High SA' and 'Low SA'
328 is around 64.4 pixels. Thus, trials with position error less than e' are labelled as high
329 SA, and the remaining trials are labelled as low SA. The SA score can then be
330 computed using Equation (1).

331

332 Second, the position errors and the angular errors are clustered using GMM. **Figure 9**
333 shows the two-dimensional plot of the clustering result.

334

335 [Figure 9 near here]

336

337 Each data point in **Figure 9** is represented by $(e_{position}, e_{angular})$. The blue dots
338 (498/700 trials, 71.1%) and green dots represent trials labelled as high and low SA by
339 the GMM, respectively.

340 4.2.3. Results of the SAGAT

341 The descriptive statistics of the SAGAT scores was illustrated in **Table 4**.

342

343 [Table 4 near here]

344

345 The results showed that the SAGAT scores were significantly different between
346 different task complexity, $F(2, 54) = 46.680, p < 0.001, \eta_p^2 = 0.634, Power = 1.000,$
347 sphericity assumed ($\chi^2 = 1.861, p = 0.394$), confirming **H₂**. Furthermore, the pairwise
348 comparison reveals that the SAGAT scores of the position identification task ($M =$
349 $0.649, SEM = 0.034, p < 0.001$) and the heading projection task ($M = 0.711, SEM =$
350 $0.024, p < 0.001$) are significantly lower than that of the callsign association task ($M =$
351 $0.927, SEM = 0.014$). No significant difference is observed between the position
352 identification and the heading projection tasks ($p = 0.186$).

353 **4.3. Eye-tracking**

354 **4.3.1. Fixations and saccadic velocity**

355 For fixations, we analysed the total duration and the number of whole fixations (**Table**
356 **5**). Whole fixation is defined as the fixations starting/ending within the TOI interval and
357 must be preceded/succeeded by a saccade [37]. Both metrics violated the sphericity
358 assumption (Total duration: $\chi^2 = 7.998, p = 0.018, \varepsilon = 0.832$; Number: $\chi^2 =$
359 $0.999, p = 0.007, \varepsilon = 0.794$), so the Huynh-Feldt corrections were adopted. There
360 were statistically significant main effects of task complexity on the total duration
361 ($F(1.663, 44.905) = 16.502, p < 0.001, \eta_p^2 = 0.379, Power = 0.744$) and the
362 number of whole fixations ($F(1.587, 42.854) = 24.325, p < 0.001, \eta_p^2 =$
363 $0.474, Power = 0.942$), confirming **H_{3a}** and **H_{3b}**.

364

365 [Table 5 near here]

366 [Table 6 near here]

367

368 The pairwise comparison (**Table 6**) shows that the total duration and the number of
369 whole fixations were significantly different between ‘Callsign & Position’ and ‘Callsign
370 & Heading’. The difference between the position identification and the heading
371 projection tasks for the total duration of whole fixations was statistically nonsignificant.

372

373 For saccades, we further analysed the maximum and average peak velocity of saccades.

374 The results are also listed in **Table 5** and **Table 6**. Both the maximum and average peak

375 velocity of saccades violated the sphericity assumption (Maximum: $\chi^2 = 13.345, p =$

376 $0.001, \varepsilon = 0.714$; Average: $\chi^2 = 8.054, p = 0.018, \varepsilon = 0.830$). Therefore,

377 Greenhouse-Geisser correction and Huynh-Feldt correction was applied to the

378 maximum and the average peak velocity of saccades, respectively. However, no
379 statistically significant main effects were observed for both the maximum
380 ($F(1.427, 38.531) = 0.808, p = 0.415, \eta_p^2 = 0.029, Power = 0.435$) and the average
381 peak velocity of saccades ($F(1.661, 44.843) = 0.968, p = 0.374, \eta_p^2 =$
382 $0.035, Power = 0.406$).

383 4.3.2. Pupil diameter

384 The pupil diameter is estimated by multiplying the image captured by the eye-tracker
385 with a scaling factor, a built-in function of Tobii Pro Lab software. The pupil
386 foreshortening error (PFE), when eyes rotate away from the camera, was mitigated by
387 the algorithm of the Tobii Pro Lab via pupil size correction.

388

389 [Table 7 near here]

390 [Table 8 near here]

391

392 **Table 7** and **Table 8** shows the descriptive statistics and the results of the one-way
393 ANOVA on pupil diameter. The sphericity assumption was not violated ($\chi^2 =$
394 $0.545, p = 0.761$). The mean pupil diameter was not significantly affected by the task
395 requirement, $F(2,54) = 2.491, p = 0.092, \eta_p^2 = 0.084, Power = 0.221$. The mean
396 pupil size (SEM) for the callsign association task, position identification task, and
397 heading projection task were 4.074 (0.112) mm, 4.096 (0.119) mm, and 4.115 (0.120)
398 mm, respectively. The pairwise comparison showed the difference in mean pupil
399 diameter between each pair of tasks is statistically nonsignificant, with p -value for task
400 pairs Callsign & Position, Callsign & Heading, and Position & Heading of 0.713, 0.137,
401 and 0.844, respectively. H_{3c} is thus rejected.

402 5. Discussions

403 5.1. *Gaining SA from fixations*

404 The results revealed that the total duration and the number of whole fixations differ
405 significantly between tasks with different complexity. The pairwise comparison
406 evidenced that participants spent a longer fixation duration and had more fixations in
407 the position identification and heading projection tasks. Indeed, when the task
408 complexity rises, human operators should be more attentive to certain positions on the
409 radar screen to cope with the task requirement. In the callsign association task,
410 participants are given four choices of the callsign, and they are required to match the
411 callsign with the position. In addition to the limited number of choices, they merely
412 need to remember the position and movement of the aircraft roughly. Therefore, this
413 task is relatively less complex and demands shorter and fewer fixations. However, with
414 the gradual increase in complexity, the position identification task requires the
415 participants to remember the position of the flights precisely and respond to one of the
416 queried flights. The heightened complexity demands participants to fixate longer and
417 more frequently during the monitoring stage to ensure the preciseness of their response.
418 In particular, the position of the flight changes over time, so they have to pay an
419 additional number of fixations to trace the latest positions. Furthermore, the heading
420 projection task is even more complex in that participants have to trace the angular
421 movements of the aircraft. It demands an even longer fixation duration to
422 simultaneously gain positional and angular information from the radar screen to
423 maintain their SA. However, the difference between the position identification task and
424 this task is nonsignificant, as is the similar nature of obtaining precise information from
425 the radar screen. Summing up, fixations can provide a clue to understanding the SA
426 development processes through visual attention distribution.

427 **5.2. Cognitive workload and SA**

428 While participants fixate on maintaining their SA, we obtained a lower SAGAT score in
429 more complex tasks that generate more and longer fixations, which does not align with
430 the literature. Indeed, several studies suggested that fixations are related to a higher
431 SAGAT score [26,27,38]. Nevertheless, it could be attributed to the impacts of
432 cognitive workload on SA. In our study, the results of the NASA-TLX indicate that the
433 cognitive workload of the heading projection task is significantly higher than the other
434 two tasks. At the same time, the position identification task is also significantly higher
435 than the callsign association task. With the increasing complexity demand, participants
436 may perceive a higher cognitive workload. Under a high cognitive workload, human
437 performance usually deteriorates such that they are not as capable of perceiving
438 environmental elements as in a low cognitive workload state. It can also be explained by
439 the ‘look-but-not-see’ phenomenon [39]. Participants might merely fixate at a certain
440 position on the screen but did not perceive any environmental element changes. Thus,
441 the reduced human performance may impede participants from gaining SA and explains
442 why the increase in fixation count and duration does not show a higher SA.

443 **5.3. Pupil size**

444 However, the result revealed that the task complexity did not significantly influence the
445 mean pupil diameter. It echoes with Zhang, Yang [25] that many confounding factors
446 affect pupil size. It could be due to individual differences between the subjects. Indeed,
447 the ‘original’ pupil size in the rest condition varies across individuals. A single
448 ‘original’ pupil size cannot fit all. Furthermore, while the relationship between pupil
449 size and cognitive workload was investigated in the past century, no consistent
450 conclusions were drawn across studies [28,40,41]. Therefore, our study concluded that
451 the pupil size does not vary with the task requirement in radar screen monitoring tasks

452 of the air traffic control environment. Hence, pupil size is unlikely to be an effective
453 predictor of SA and cognitive workload.

454 **5.4. Implications**

455 This study reveals that the human operators mainly use fixations to perceive the flight's
456 callsign, position, and heading displayed on the radar screen. Human operators attempt
457 to fixate more and longer to maintain their SA in more complex tasks. Therefore,
458 fixation-related metrics can be used to recognise the visual attention that human
459 operators of radar monitoring tasks were allocated to maintain their SA. However,
460 results indicate that task complexity simultaneously raises the cognitive workload as a
461 side effect. The increase in cognitive workload corresponds to a drop in SA with the
462 increased task complexity. It is because they are less capable of perceiving the
463 environment even though they fixate longer to attempt the information. It implies that
464 the cognitive workload shall be controlled at an acceptable level that shall not induce
465 adverse effects on SA. Furthermore, past studies may merely consider a rule of thumb
466 in defining criteria like distinguishing high and low SA. However, our study proposes
467 that using GMM to cluster SA response makes the criterion more objective, facilitating
468 a more objective classification criterion that can be adopted with trust in the application
469 context.

470 **6. Concluding remarks**

471 Overall, the current study investigates the gaze behaviours, situational awareness (SA),
472 and cognitive workload of air traffic controllers during different radar screen monitoring
473 tasks with varying task complexity. Our study employed GMM to enhance the
474 objectiveness in defining high and low SA. Also, empirical results suggest that the
475 cognitive workload and SA significantly differ between task complexity. Furthermore,

476 fixation-related metrics may signify the changes in situational awareness under varying
477 task complexity. However, pupil size is unlikely an effective predictor of SA and
478 cognitive workload.

479

480 The current study is also subject to several limitations. First, we only considered tasks
481 on SA Level 1. These perceiving tasks focus on ATC's primary element extraction tasks
482 and are fundamental for ATCOs to maintain their SA. Future studies can evaluate the
483 effects on SA Level 2 or Level 3 tasks like conflict resolution and motion projection.
484 Furthermore, our participants are students and staff with aviation and ATC backgrounds
485 instead of ATC practitioners. While the tasks are general and the participants have an
486 appropriate level of expertise to accomplish the tasks, these data are also critical as
487 relatively naïve participants could reveal the “real” workload and difficulty of the tasks.
488 However, future studies may also compare if there are any differences between
489 practitioners and non-practitioner participants with relevant knowledge, as well as
490 generalising the results to real-life ATC work with ATCOs samples. Ultimately, the
491 current study can serve as a cornerstone to support further real-time human performance
492 monitoring in the ATC environment to secure flight safety.

493

494 **Declarations of interest:** The authors declare that they have no known competing
495 financial interests or personal relationships that could have appeared to influence the
496 work reported in this paper.

497

498 **References**

- 499 [1] Durso FT, Manning CA. Air Traffic Control. *Reviews of Human Factors and*
500 *Ergonomics*. 2008;4(1):195-244.
- 501 [2] Loft S, Sanderson P, Neal A, et al. Modeling and Predicting Mental Workload in
502 *En Route Air Traffic Control: Critical Review and Broader Implications*. *Human*
503 *Factors*. 2007;49(3):376-399.
- 504 [3] Yeung TC, Yiu CY, Ng KKH, et al., editors. Airside terminal traffic flow
505 problem formulation under extreme weather: A case study in the Hong Kong
506 International Airport. 2022 Joint 12th International Conference on Soft
507 Computing and Intelligent Systems and 23rd International Symposium on
508 Advanced Intelligent Systems (SCIS&ISIS); 2022.
- 509 [4] Li Q, Ng KKH, Yu SCM, et al. Recognising situation awareness associated with
510 different workloads using EEG and eye-tracking features in air traffic control
511 tasks. *Knowledge-Based Systems*. 2023 2023/01/25/;260:110179.
- 512 [5] Li Q, Leung HC, Ho MH, et al. The Effects of Aeronautical Decision-Making
513 Models on Student Pilots' Situational Awareness and Cognitive Workload in
514 Simulated Non-Normal Flight Deck Environment. *The International Journal of*
515 *Aerospace Psychology*. 2023 2023/07/03;33(3):197-213.
- 516 [6] Chen M-L, Lu S-Y, Mao IF. Subjective symptoms and physiological measures
517 of fatigue in air traffic controllers. *International Journal of Industrial*
518 *Ergonomics*. 2019 2019/03/01/;70:1-8.
- 519 [7] Yiu CY, Ng KKH, Yu SCM, et al. Sustaining aviation workforce after the
520 pandemic: Evidence from Hong Kong aviation students toward skills,
521 specialised training, and career prospects through a mixed-method approach.
522 *Transport Policy*. 2022;128:179-192.
- 523 [8] Yiu CY, Ng KKH, Chan FTS, et al. Vaccines, associated risk and air transport
524 industry post-COVID-19: A structural equation modelling-based empirical study
525 in Hong Kong. *Research in Transportation Business & Management*. 2023
526 2023/10/01/;50:101038.
- 527 [9] Yiu CY, Ng KKH, Li X, et al. Towards safe and collaborative aerodrome
528 operations: Assessing shared situational awareness for adverse weather detection
529 with EEG-enabled Bayesian neural networks. *Advanced Engineering*
530 *Informatics*. 2022;53:101698.
- 531 [10] Rowen A, Grabowski M, Rancy J-P. Moving and improving in safety-critical
532 systems: Impacts of head-mounted displays on operator mobility, performance,
533 and situation awareness. *International Journal of Human-Computer Studies*.
534 2021 2021/06/01/;150:102606.
- 535 [11] Yiu CY, Ng KKH, Wong KY, et al. On the significance of using event-related
536 potentials and gaze behaviours to evaluate inflight emergency training
537 performance. *The 26th ATRS World Conference; Kobe, Japan2023*.
- 538 [12] Li Q, Ng KKH, Fan Z, et al. A human-centred approach based on functional
539 near-infrared spectroscopy for adaptive decision-making in the air traffic control
540 environment: A case study. *Advanced Engineering Informatics*. 2021
541 2021/08/01/;49:101325.
- 542 [13] Yiu CY, Ng KKH, Li Q, et al., editors. How do the levels of automation in flight
543 operations affect pilots' cognitive workload, reaction time, and EEG brain
544 waves in cruising flights? 2024 IEEE 4th International Conference on Human-
545 Machine Systems (ICHMS); 2024 15-17 May 2024.

- 546 [14] Li Q, Chen C-H, Ng KKH, et al. Single-pilot operations in commercial flight:
547 Effects on neural activity and visual behaviour under abnormalities and
548 emergencies. *Chinese Journal of Aeronautics*. 2024 2024/04/12/.
- 549 [15] Liang N, Yang J, Yu D, et al. Using eye-tracking to investigate the effects of
550 pre-takeover visual engagement on situation awareness during automated
551 driving. *Accident Analysis & Prevention*. 2021 2021/07/01/;157:106143.
- 552 [16] Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems.
553 *Human Factors*. 1995;37(1):32-64.
- 554 [17] Wilson GF, Caldwell JA, Russell CA. Performance and Psychophysiological
555 Measures of Fatigue Effects on Aviation Related Tasks of Varying Difficulty.
556 *The International Journal of Aviation Psychology*. 2007 2007/04/17;17(2):219-
557 247.
- 558 [18] Di Stasi LL, Diaz-Piedra C, Suárez J, et al. Task complexity modulates pilot
559 electroencephalographic activity during real flights. *Psychophysiology*. 2015
560 2015/07/01;52(7):951-956.
- 561 [19] Li P-C, Wang Y-X, Chen J-H, et al. An experimental study on the effects of task
562 complexity and knowledge and experience level on SA, TSA and workload.
563 *Nuclear Engineering and Design*. 2021 2021/05/01/;376:111112.
- 564 [20] Nguyen T, Lim CP, Nguyen ND, et al. A Review of Situation Awareness
565 Assessment Approaches in Aviation Environments. *IEEE Systems Journal*.
566 2019;13(3):3590-3603.
- 567 [21] Endsley MR, editor Situation awareness global assessment technique (SAGAT).
568 *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*;
569 1988 23-27 May 1988.
- 570 [22] Durso FT, Hackworth CA, Truitt TR, et al. Situation awareness as a predictor of
571 performance for en route air traffic controllers. *Air Traffic Control Quarterly*.
572 1998;6(1):1-20.
- 573 [23] Paige Bacon L, Strybel TZ. Assessment of the validity and intrusiveness of
574 online-probe questions for situation awareness in a simulated air-traffic-
575 management task with student air-traffic controllers. *Safety Science*. 2013
576 2013/07/01/;56:89-95.
- 577 [24] Lyu M, Li F, Xu G, et al. Leveraging eye-tracking technologies to promote
578 aviation safety- A review of key aspects, challenges, and future perspectives.
579 *Safety Science*. 2023 2023/12/01/;168:106295.
- 580 [25] Zhang T, Yang J, Liang N, et al. Physiological Measurements of Situation
581 Awareness: A Systematic Review. *Human Factors*.
582 2020;0(0):0018720820969071.
- 583 [26] Hasanzadeh S, Esmaeili B, Dodd MD. Examining the Relationship between
584 Construction Workers' Visual Attention and Situation Awareness under Fall and
585 Tripping Hazard Conditions: Using Mobile Eye Tracking. *Journal of*
586 *Construction Engineering and Management*. 2018;144(7):04018060.
- 587 [27] Moore K, Gugerty L. Development of a Novel Measure of Situation Awareness:
588 The Case for Eye Movement Analysis. *Proceedings of the Human Factors and*
589 *Ergonomics Society Annual Meeting*. 2010;54(19):1650-1654.
- 590 [28] Marquart G, Cabrall C, de Winter J. Review of Eye-related Measures of Drivers'
591 Mental Workload. *Procedia Manufacturing*. 2015 2015/01/01/;3:2854-2861.
- 592 [29] Dreyer D, Oberhauser M, Bandow D. HUD symbology evaluation in a virtual
593 reality flight simulation. *Proceedings of the International Conference on*
594 *Human-Computer Interaction in Aerospace*; Santa Clara, California: Association
595 for Computing Machinery; 2014. p. Article 9.

- 596 [30] Paletta L, Dini A, Murko C, et al. Towards Real-time Probabilistic Evaluation of
597 Situation Awareness from Human Gaze in Human-Robot Interaction.
598 Proceedings of the Companion of the 2017 ACM/IEEE International Conference
599 on Human-Robot Interaction; Vienna, Austria: Association for Computing
600 Machinery; 2017. p. 247–248.
- 601 [31] Zhou F, Yang XJ, Winter JCFd. Using Eye-Tracking Data to Predict Situation
602 Awareness in Real Time During Takeover Transitions in Conditionally
603 Automated Driving. IEEE Transactions on Intelligent Transportation Systems.
604 2021:1-12.
- 605 [32] de Winter JCF, Eisma YB, Cabrall CDD, et al. Situation awareness based on eye
606 movements in relation to the task environment. Cognition, Technology & Work.
607 2019 2019/02/01;21(1):99-111.
- 608 [33] Yiu CY, Ng KKH, Lee C-H, et al. A Digital Twin-Based Platform towards
609 Intelligent Automation with Virtual Counterparts of Flight and Air Traffic
610 Control Operations. Applied Sciences. 2021;11(22):10923.
- 611 [34] Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index):
612 Results of empirical and theoretical research. In: Meshkati PAHaN, editor.
613 Human Mental Workload. Amsterdam: North Holland Press; 1988.
- 614 [35] Kästle JL, Anvari B, Krol J, et al. Correlation between Situational Awareness
615 and EEG signals. Neurocomputing. 2021 2021/04/07;432:70-79.
- 616 [36] Hair JF. Multivariate data analysis. 2009.
- 617 [37] Tobii Pro Learning Center. Understanding Pro Lab's eye tracking metrics 2021.
618 Available from: [https://www.tobiipro.com/learn-and-support/learn/steps-in-an-
619 eye-tracking-study/data/understanding-Pro-Labs-eye-tracking-metrics/](https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/data/understanding-Pro-Labs-eye-tracking-metrics/)
- 620 [38] Yuan X, Ng KKH, Li Q, et al. Exploring the Human-Centric Interaction
621 Paradigm: Augmented Reality-Assisted Head-Up Display Design for
622 Collaborative Human-Machine Interface in Cockpit. Advanced Engineering
623 Informatics. 2024 2024/10/01;62:102656.
- 624 [39] Li Q, Ng KKH, Yiu CY, et al. Securing air transportation safety through
625 identifying pilot's risky VFR flying behaviours: An EEG-based
626 neurophysiological modelling using machine learning algorithms. Reliability
627 Engineering & System Safety. 2023;238:109449.
- 628 [40] Charles RL, Nixon J. Measuring mental workload using physiological measures:
629 A systematic review. Applied Ergonomics. 2019 2019/01/01;74:221-232.
- 630 [41] Bradshaw J. Pupil Size as a Measure of Arousal during Information Processing.
631 Nature. 1967 1967/11/01;216(5114):515-516.

632

633

634 **Appendix.** Skewness and kurtosis of the measures

Measure	Skewness			Kurtosis		
	Callsign	Position	Heading	Callsign	Position	Heading
<i>NASA-TLX</i>						
Mental Demand	-0.405	-0.638	-0.21	-0.356	-0.187	-0.434
Temporal Demand	0.016	0.003	0.014	0.575	-0.558	-0.883
Performance (-ve)	0.657	0.417	0.389	-0.34	-0.159	-0.667
Effort	0.475	0.111	-0.149	0.93	-0.063	-0.597
Frustration	-0.3	-0.599	-0.62	-0.555	-0.155	-0.264
Average	-0.717	-0.143	-0.122	0.737	-0.563	-0.826
SAGAT	-1.75	-1.008	0.008	2.276	-0.188	-1.278
<i>Eye-tracking</i>						
Total duration of whole fixations	-0.487	-0.498	-0.895	-0.651	-0.733	0.422
Number of whole fixations	-0.528	-0.466	-0.941	-0.612	-0.826	0.792
Peak velocity of saccades (Max.)	-0.958	-0.729	-0.589	0.718	0.76	1.121
Peak velocity of saccades (Avg.)	-0.387	-0.17	-0.175	-0.504	-0.031	0.618
Pupil Diameter	-0.276	-0.088	-0.03	-0.332	-0.405	-0.344

635 Note: Callsign = Callsign association, Heading = Heading projection, Position =

636 Position identification

637

638 **Titles of Tables**

639 **Table 1.** Descriptive statistics of the cognitive workload measures.

640 **Table 2.** Results of one-way repeated measures ANOVA on cognitive workload
641 measures.

642 **Table 3.** Pairwise comparison for the one-way repeated measures ANOVA on cognitive
643 workload measures.

644 **Table 4.** Descriptive statistics on SAGAT.

645 **Table 5.** Descriptive statistics on fixation and saccade metrics.

646 **Table 6.** Pairwise comparison for the one-way repeated measures ANOVA on fixation
647 and saccade metrics.

648 **Table 7.** Descriptive statistics and results of the one-way repeated measures ANOVA
649 on pupil diameter.

650 **Table 8.** Pairwise comparison for the one-way repeated measures ANOVA on fixation-
651 related metrics.

652

653 **Tables**

654 **Table 1.** Descriptive statistics of the cognitive workload measures.

Dimension	<i>M</i> (SEM)		
	Callsign	Position	Heading
Mental Demand	8.61 (0.695)	11.11 (0.530)	14.79 (0.607)
Temporal Demand	7.75 (0.710)	10.07 (0.738)	13.43 (0.827)
Performance (-ve)	5.11 (0.683)	6.93 (0.577)	10.5 (0.607)
Effort	9.79 (0.772)	11.61 (0.629)	14.86 (0.601)
Frustration	7.21 (0.756)	9.61 (0.797)	12.25 (0.978)
Average	7.69 (0.471)	9.86 (0.471)	13.16 (0.543)

655 Note: -ve = reversely coded, a higher value implies less performance, Callsign =

656 Callsign association, Heading = Heading projection, Position = Position identification

657

658 **Table 2.** Results of one-way repeated measures ANOVA on cognitive workload

659 measures.

Dimension	F	<i>p</i>	η_p^2	Power
Mental Demand	$F(1.420, 38.331) = 41.21$	< 0.001*	0.604	0.996
Temporal Demand	$F(1.492, 40.272) = 37.397$	< 0.001*	0.850	0.994
Performance (-ve)	$F(1.734, 46.810) = 35.352$	< 0.001 [#]	0.567	0.998
Effort	$F(1.680, 45.353) = 37.933$	< 0.001 [#]	0.584	0.998
Frustration	$F(1.449, 39.118) = 26.746$	< 0.001*	0.498	0.946
Average	$F(1.580, 42.661) = 64.781$	< 0.001 [#]	0.706	1.000

660 Note: -ve = reversely coded, a higher value implies less performance, * = Greenhouse-

661 Geisser correction applied, [#] = Huynh-Feldt correction applied

662

663 **Table 3.** Pairwise comparison for the one-way repeated measures ANOVA on cognitive
 664 workload measures.

Dimension	Callsign-Position	Position-Heading	Callsign-Heading
Mental Demand	$p < 0.001$	$p < 0.001$	$p < 0.001$
Temporal Demand	$p < 0.001$	$p < 0.001$	$p < 0.001$
Performance (-ve)	$p = 0.007$	$p < 0.001$	$p < 0.001$
Effort	$p = 0.020$	$p < 0.001$	$p < 0.001$
Frustration	$p = 0.001$	$p < 0.001$	$p < 0.001$
Average	$p < 0.001$	$p < 0.001$	$p < 0.001$

665 Note: -ve = reversely coded, a higher value implies less performance, Callsign =
 666 Callsign association, Heading = Heading projection, Position = Position identification
 667

668 **Table 4.** Descriptive statistics on SAGAT.

Dependent measure	<i>M</i> (SEM)		
	Callsign	Position	Heading
SAGAT	0.927 (0.014)	0.649 (0.034)	0.711 (0.024)

669 Note: Callsign = Callsign association, Heading = Heading projection, Position =

670 Position identification, Units = Percentage

671

672 **Table 5.** Descriptive statistics on fixation and saccade metrics.

Dependent measure	<i>M</i> (SEM)		
	Callsign	Position	Heading
Total duration of whole fixations	2358.029	2611.097	2845.749
(s)	(190.414)	(206.5008)	(191.286)
Number of whole fixations	8.987 (0.723)	10.179 (0.817)	11.357 (0.737)
Peak velocity of saccades (Max.)	321.584	323.280	329.640
	(3.520)	(3.291)	(3.002)
Peak velocity of saccades (Avg.)	161.637	156.534	159.244
	(1.549)	(1.368)	(1.200)

673 Note: Avg. = Average, Callsign = Callsign association, Heading = Heading projection,

674 Max. = Maximum, Position = Position identification, s = seconds

675

676 **Table 6.** Pairwise comparison for the one-way repeated measures ANOVA on fixation
 677 and saccade metrics.

Dependent measure	Callsign- Position	Position- Heading	Callsign- Heading
Total duration of whole fixations	$p = 0.001$	$p < 0.001$	$p = 0.062$
Number of whole fixations	$p < 0.001$	$p < 0.001$	$p = 0.014$
Peak velocity of saccades (Max.)	$p = 1.000$	$p = 0.999$	$p = 1.000$
Peak velocity of saccades (Avg.)	$p = 0.269$	$p = 1.000$	$p = 1.000$

678 Note: Avg. = Average, Callsign = Callsign association, Heading = Heading projection,
 679 Max. = Maximum, Position = Position identification, s = seconds

680

681 **Table 7.** Descriptive statistics and results of the one-way repeated measures ANOVA
 682 on pupil diameter.

Dependent measure	<i>M</i> (SEM)			<i>F</i>	<i>p</i>	η_p^2
	Callsign	Position	Heading			
Pupil Diameter (mm)	4.074 (0.112)	4.096 (0.119)	4.115 (0.120)	<i>F</i> (2,54) = 2.491	0.092 ⁺	0.084

683 Note: ⁺ = Sphericity assumed, Callsign = Callsign association, Heading = Heading

684 projection, mm = millimetres, Position = Position identification

685

686 **Table 8.** Pairwise comparison for the one-way repeated measures ANOVA on fixation-
 687 related metrics.

Dependent measure	Callsign-Position	Position-Heading	Callsign-Heading
Pupil diameter (mm)	$p = 0.713$	$p = 0.844$	$p = 0.137$

688 Note: Callsign = Callsign association, Heading = Heading projection, mm =
 689 millimetres, Position = Position identification
 690

691 **Titles of Figures**

692 **Figure 1.** The experimental setting.

693 **Figure 2.** Monitoring period for all tasks.

694 **Figure 3.** Callsign association task.

695 **Figure 4.** Position identification task.

696 **Figure 5.** Heading projection task.

697 **Figure 6.** Experimental procedure. Note: NASA-TLX = National Aeronautics and
698 Space Administration Task Load Index.

699 **Figure 7.** Data analysis. Note: ° = Degrees, ANOVA = Analysis of Variance, GMM =
700 Gaussian Mixture Model, NASA-TLX = National Aeronautics and Space
701 Administration Task Load Index, PDF = Probability Density Function, SA = Situational
702 Awareness, SAGAT = Situational Awareness Global Assessment Technique, TOI =
703 Time of Interest.

704 **Figure 8.** Probability density function fitted by the GMM for the position identification
705 task. Note: GMM = Gaussian Mixture Model. The full colour version of this figure is
706 available online.

707 **Figure 9.** Result of the clustering via GMM for the heading projection task. Note: ° =
708 Degrees, GMM = Gaussian Mixture Model. The full colour version of this figure is
709 available online.

710