The Aeronautical Journal (2024), **128**, pp. 2191–2217
doi:10.1017/aer.2024.35

RESEARCH ARTICLE

Pilot performance during simulated point and boundary avoidance tracking tasks

Q. Xia[1,2], D. Marchesoli[1], P. Masarati[1] and M. Liu[2]

[1]Department of Aerospace Science and Technology, Politecnico di Milano, Milano, 20156, Italy
[2]School of Aeronautic Science and Engineering, Beihang University, Beijing, 100191, China
Corresponding author: P. Masarati; Email: pierangelo.masarati@polimi.it

Received: 2 May 2023; **Revised:** 4 January 2024; **Accepted:** 19 March 2024

Keywords: aircraft-pilot coupling; rotorcraft-pilot coupling; boundary avoidance tracking; piloting strategies

Abstract
Helicopters are used in complex and harsh operational environments, such as search and rescue missions and firefighting, that require operating in ground proximity, tracking targets while avoiding impacting obstacles, namely a combination of point tracking (positive) and boundary avoidance (negative) objectives. A simulation task representing simplified helicopter dynamics is used to investigate point tracking and boundary avoidance tasks. The variance and regression analysis are used to study the effects of task conditions on participants' tracking errors and input aggression. The overall tracking error shows a negative correlation with input aggression. The participants tend to have higher input aggression and lower tracking error near the boundaries, exposing the switching of manipulation input strategies under different task conditions. It also suggests a potential way of designing simulation tasks for human operators manipulating helicopters and a trigger for investigating pilots' biodynamic feedthrough.

List of acronyms

ANOVA	analysis of variance
APC	aircraft-pilot coupling
BAT	boundary avoidance tracking
BDFT	biodynamic feedthrough
GUI	graphical user interface
MAE	mean average error
PAO	pilot assisted oscillations
PIO	pilot induced oscillations
RMS	root mean square
RMSE	root mean square error
RPC	rotorcraft-pilot coupling
GPR	Gaussian process regression

1.0 Introduction

1.1 Aircraft and rotorcraft pilot coupling

Current high-performance aircraft and rotorcraft have been developed to meet the increasing demands of various missions. During design, testing, and operations, engineers and pilots must anticipate and manage unfavourable occurrences known as aircraft-pilot couplings (APCs), or rotorcraft-pilot couplings (RPCs) [1] in the fixed- and rotary-wing contexts, respectively. APC and RPC phenomena emerge from the undesired and unexpected coupling between the pilot and the vehicle. They can lead to instabilities of both oscillatory and non-oscillatory nature, degraded handling qualities and increasing structural strength requirements, potentially resulting in catastrophic accidents. APC and RPC 'events are rare, unexpected, and unintended' [2], but the frequency of their insurgence can be expected to increase and manifest itself in different and unforeseen forms as new types of air vehicles are introduced [3].

Before 1995, the aircraft and rotorcraft communities referred to APC and RPC events as pilot-induced oscillations (PIO) and pilot-assisted oscillations (PAO) [4, 5], whose distinction carried an underlying, implicit emphasis on whether the interaction was with rigid-body or aeroelastic vehicle dynamics [1]. Although such denominations appear to put inappropriate and excessive emphasis on the role of pilots as the primary cause of APCs and RPCs, whereas the culprit of the phenomenon lies in the proneness of the vehicle to adverse interaction with the pilot [2], they remain of general use. In this context, in accordance with [1] and further highlighting their underlying genesis and nature, we choose to interpret PIOs as 'rigid-body A/RPCs' or *control-motivated* A/RPCs, and PAOs as 'aeroelastic A/RPCs' or *control-unrelated* A/RPCs.

1.2 Boundary avoidance tracking

Boundary-avoidance tracking (BAT) is a pilot-task model proposed by Gray [6, 7], which stems from the consideration that in the process of performing flight tasks, pilots not only need to complete the task of 'maintaining specific conditions' (point tracking, a 'positive' objective) but also typically need to 'avoid certain conditions' (boundary avoidance, a 'negative' objective), such as clearance from obstacles, or staying away from dangerous operational conditions, like stall for fixed-wing aircraft or vortex-ring susceptible envelope for helicopters. Researchers believe that boundary avoidance behaviour has a strong correlation with the critical phenomenon of PIO, as defined in Section 1.1, which triggers the pilot's input change from minimal, gentle control to rapid, exaggerated commands, often in phase opposition [6, 7]. The usual point-tracking models, focusing on the positive goal of the pilot, usually cannot correctly describe the onset of the corresponding PIO phenomenon, while boundary avoidance tracking can capture the struggle between the positive and negative objectives of the pilot.

BAT is used across many domains including workload/cognitive load measurement [8, 9], pilot model investigation [10–12], flight safety [13, 14], handling qualities [15], situational awareness [16], etc. For the present study, a task inspired by the BAT concept was designed, where point tracking and boundary avoidance tracking were simultaneously employed to investigate how participants make trade-offs between the two control strategies.

1.3 Pilot's performance

During task execution, the situational awareness of the pilot in the human-machine system has a significant impact on task performance. A generally agreed definition of situational awareness is divided into three levels: perception, comprehension and projection, each playing a different role in the system [17]. The measurement of situational awareness can be divided into direct and indirect methods [18], where the former includes Situation Awareness Global Assessment Technique (SAGAT) [19], Situation Present Assessment Method (SPAM) [20] and other query methods, and the latter includes process measurement (eye movements, communication, etc.) [21, 22] and performance measurement (reaction time, error, etc.) [23, 24]. The task design of this study includes explicit tracking tasks and direct data acquisition. Therefore, point-tracking error and input aggression were used to evaluate the performance of participants, reflecting their situational awareness.

1.4 Research purpose

This study aims to discover the relationship between pilots' performance and input strategy when simultaneously pursuing contradictory task goals, namely point tracking and boundary avoidance.

This study is based on a simple hardware flight simulator. Flight simulation plays an important role in areas such as human-machine interaction, pilot modelling and situational awareness. Lu and Jump [25]

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

established and determined the pilot model and parameters under the BAT task using flight simulation tasks, while Feng et al. [26–28] used flight simulation tasks to determine the relationship between task design, situational awareness of subjects and work performance. Zanoni et al. used flight simulation to study the biomechanical feedthrough (BDFT) of the upper limbs of pilots and the effect of the task on pilots' muscular activation [29–31].

An important element of the task design of this study lies in its randomness. Several previous studies about boundary avoidance used periodic tasks [11, 12], which would cause participants to learn the regularity of the task goals and predict their trajectory, thus affecting the tracking performance and objectivity of the experiment and model fitting.

BDFT provides an insight into topics such as PAO caused by the passive, i.e., involuntary, response of the pilot. The task and interface design presented in this study are being used also for research in this area [32]. The effect of BAT on BDFT will be the objective of future research.

2.0 Objective and approach

This research attempts to answer these questions:

- RQ1: Will pilots use different input strategies to cope with different task conditions (namely, point tracking and boundary avoidance)?
- RQ2: What will the performance of the pilots be under different task conditions?
- RQ3: Is there a correlation between pilots' input strategy and the corresponding performance?

Under the guidance of these questions, this research investigates the pilots' response to the simultaneous and contrasting goals of point tracking and boundary avoidance in a simplified simulated flight task; two main aspects are the aim of this study:

1. Pilots' tracking performance and input aggression under different task conditions regarding both point tracking and boundary avoidance tracking.
2. The possible relationship between input strategy (aggression) and task performance (tracking error).

Section 3.0 describes the methodology of this study. Section 4.0 describes the data analysis procedure. Section 5.0 presents the results of the experiments that are discussed in Section 6.0. Conclusions are finally drawn in Section 7.0.

3.0 Method

3.1 Participants

Fourteen participants volunteered and took part in the experiments. Participants were randomly recruited from undergraduate and graduate students, with an average age of M = 25.08, SD = ±1.12 years. The participants had no specific helicopter training or former experience with the task. One of them was female, 12 of them had video game experience with controllers, and 4 of them had simulated or real-life fixed-wing piloting experience. The participants were briefed about the test procedure and its objectives, without excessive details, to avoid influencing the control strategies they were going to use. They were informed that they could call the end of the experiment at any time if they felt any physical or mental discomfort. All participants signed an informed consent form prior to the experiment.

3.2 Test devices

The joystick utilised in this study was a ThrusterMaster developed by Guillemot Corporation S.A., France, shown in Fig. 1. It has two main sticks. The left one only moves in the longitudinal direction,

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

Figure 1. *The joystick used for the test.*

Figure 2. *The experiment's set-up.*

simulating the 'collective' inceptor of a helicopter. The right one moved in both longitudinal and lateral directions, simulating the 'cyclic' inceptor of a helicopter. The simulated task was developed and operated on a laptop, with an Intel Core i5-7300HQ processor, Nvidia Geforce GTX1060 6 GB VRAM graphics card, and 16 GB of RAM. The laptop's operating system was Windows 10 21H2, version 19044.1645.220403-0835. The joystick was connected to the laptop using its USB cable. Figure 2 shows the set-up of the experiment. In this study, only the cyclic stick was used.

3.3 Experiment design and procedures

3.3.1 Task design

The tasks were designed based on a helicopter tracking task and the concepts of 'point tracking' and BAT. During the task, the participants received visual information from the monitor. Two types of information were displayed on the monitor, the 'point' to be tracked and the 'boundary' to be avoided. Two

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

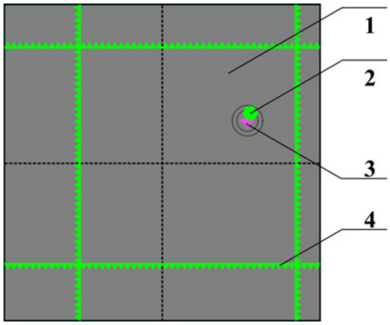

Figure 3. *GUI interface.*

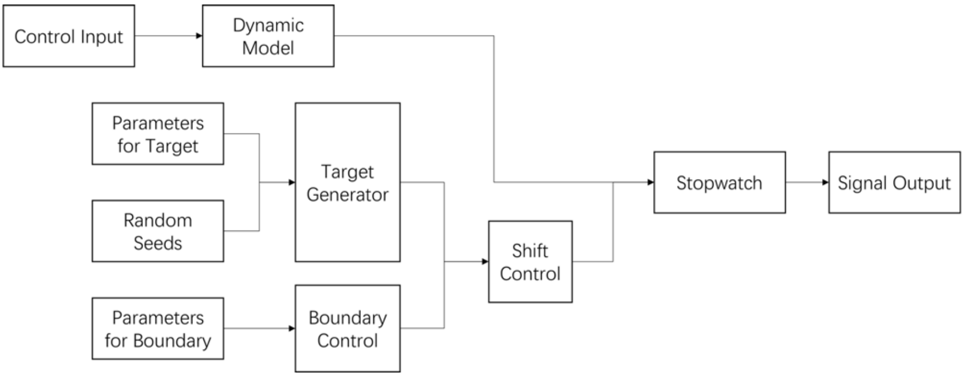

Figure 4. *Simulink module.*

types of tasks were performed: one without boundary, where participants could focus on the point tracking task; the other with boundary, where the participants were told to conduct a point tracking task (a positive objective) while avoiding the boundaries (a negative objective), such that once the boundary was touched, the task would fail immediately. As a consequence, particular emphasis was placed on avoiding the boundaries, whereas no specific reward or emphasis was associated with tracking accuracy. The same task design will be utilised also for future/ongoing research in pilots' BDFT.

3.3.2 GUI design

The GUI was designed in scalable vector graphics (SVG) format. The participants would see the interface of Fig. 3, whose main elements are:

1. A square scale, related to the cyclic stick
2. A circular indicator (in green), displaying the position of the cursor that represents the response to a cyclic input
3. A purple diamond, indicating the target that the pilots are requested to track with the above-mentioned circular indicator
4. Sawtooth boundaries, for boundary avoidance tracking tasks

The vertical and horizontal axes of the square scale correspond to the longitudinal and lateral components of the cyclic stick's motion, respectively. Among the displayed elements in Fig. 3, the element indicated with (4) and (5) would change colour according to the distance between the target and the pointer, as an indicator for the participants to adjust their controlling strategies; the element indicated

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

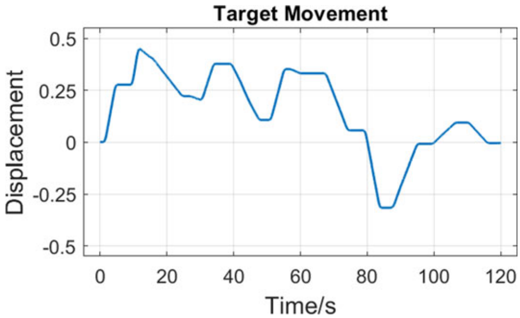

Figure 5. *Target movement.*

with (6) would also change colour if the pointer were close to that specific boundary, as a visual proximity warning.

3.3.3 Simulink module

The Simulink module of Fig. 4 was integrated into MATLAB 2022a to generate and output the signals related to target and boundary movement, collect the input signals from the joystick and filter them through the helicopter transfer function discussed in Section 3.3.4.

The blocks of the Simulink module are:

- Control Input: it receives input signals from the joystick (eventually, from the flight simulator platform).
- Dynamic Model: it simulates the dynamics of the helicopter model. For this research, an appropriately crafted second-order transfer function was used.
- Target Generator: it receives parameters and random seeds to generate the signal that controls the target movement.
- Boundary Control: it receives the parameters to control the movement of boundaries.
- Shift Control: it adds the shifting movement to the boundary and target movement.
- Stopwatch: it terminates the task when a certain condition is met (e.g., the task duration time is exceeded or the boundaries are crossed).

Several parameters control the movement of the target and boundaries. For target movement, the adjustable parameters include motion speed, direction, and duration, stance duration, maximum position and task duration. For boundary motion, the adjustable parameters include motion speed, lower limit position and type of motion pattern. Random target motion parameters were set within certain limits. Figure 5 shows an example of one-axis target movement.

Random seeds were utilised to generate reproducible random signals. By selecting a sequence of random seeds, the participants would individually experience a set of unpredictable random tasks, while tasks were consistent among all the participants.

Boundary movement patterns could be configured as 'discrete' or 'continuous'. An example of one-axis boundary movement is shown in Fig. 6. It illustrates the difference between discrete and continuous boundary motion. The transition between discrete boundary changes was always smoothed, to provide the participants with some form of warning that the boundaries were about to change.

A shift could be applied to both the target and boundary movement signals. A sinusoidal function with increasing frequency ('chirp') was added to the target and boundary movement signals. The amplitude of the sinusoidal function was parameterised and adjusted in real-time so that the boundaries would not

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

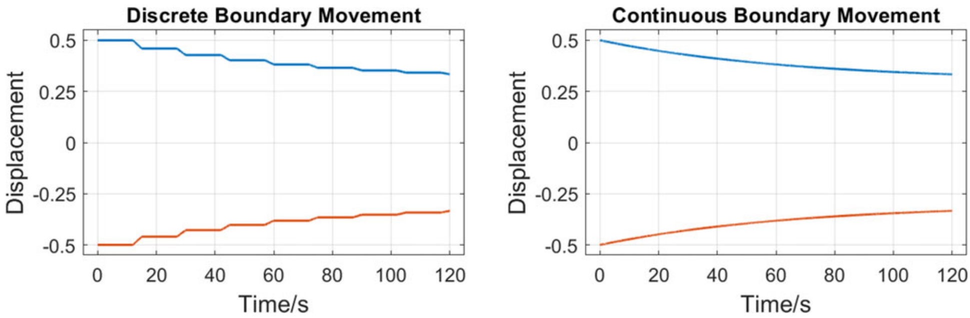

Figure 6. *"Discrete" and "continuous" boundary movement.*

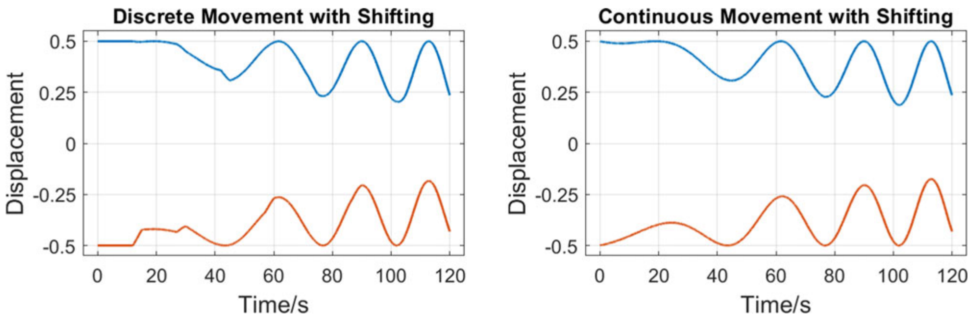

Figure 7. *Boundary movement with shifting.*

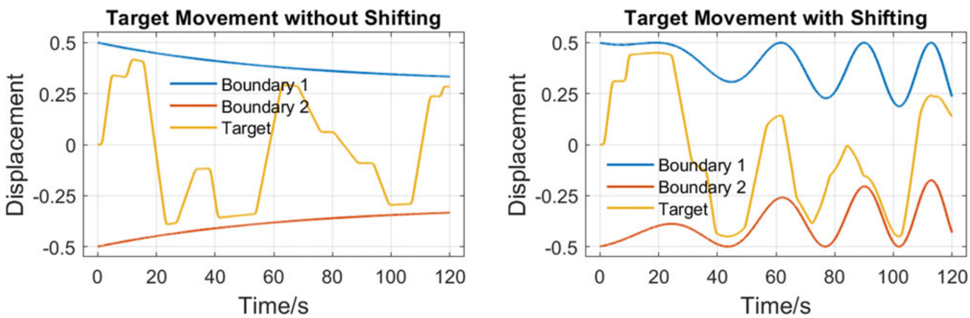

Figure 8. *Target movement with and without shifting.*

move outside of the scale. Figure 7 shows the boundary movement with shifting. And Fig. 8 shows how the target and boundary move with and without shifting.

In the definition of the tasks, care was taken to make them always attainable; no task required getting too close to a boundary or even having to cross one, to reach the target.

3.3.4 Simplified helicopter dynamics

A second-order transfer function is implemented in this study as the dynamic model of Fig. 4. The structure of this transfer function mimics the function used in [33] to describe the helicopter dynamics along the vertical axis in hover, which contains the dominant pole of the helicopter and an integrator, expressing the relationship between the control input and the displacement of the helicopter. The parameters of the transfer function have been tuned to find a trade-off between realism and feasibility/difficulty of

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

the task. No claim is made on the fidelity or even the representativeness of such a dynamic model. The resulting transfer function is

$$H(s) = \frac{0.45}{s^2 + 0.3s} \tag{1}$$

3.3.5 Task patterns

The participants were instructed to operate only the cyclic stick, as operating two sticks at the same time proved to be too difficult among research group members under the configuration for this study. Correspondingly, target and boundary movements were only applied to the cyclic stick. The participants performed a sequence of five different types of tasks, three runs for each type (except for Task 0, which the participants could repeat as many times as they wished, for familiarisation). The types of tasks are:

- Task 0: Point tracking task only, with no termination condition other than task duration. The maximum target movement was 60% of the scale's dimension. This task was for the participants to familiarise themselves with the joystick and response transfer function.
- Task 1: Point tracking task with boundary avoidance tracking, boundary movement is discrete, no shifting movement.
- Task 2: Point tracking task with boundary avoidance tracking, boundary movement is continuous, no shifting movement.
- Task 3: Point tracking task with boundary avoidance tracking, boundary movement is discrete, with shifting movement.
- Task 4: Point tracking task with boundary avoidance tracking, boundary movement is continuous, with shifting movement.

Data collected during the experiment included:

- Target movement
- Boundary movement
- Raw input signals
- Pointer movement signals (input filtered by the transfer function)

All data were collected for both the longitudinal and lateral axes of the cyclic stick.

3.3.6 Index definition

In this study, tracking error and input aggression are utilised to evaluate the performance and control strategy of the participants. The error is defined for the purpose of evaluating how well the participants performed the point-tracking task. Aggression is defined to evaluate how intensely the participants were manipulating the joystick [34]. The larger the input amplitude or frequency, the larger the aggression is. Several definitions have been proposed, which differ in the norm that is used to evaluate the rate of the control input. Lu and Jump [25] proposed to use its 1-norm; in this work, the root mean square (RMS) is used instead, following the work of Gray [35].

The indicators are defined as

$$\text{error} = \text{target} - \text{response} \tag{2}$$

$$\text{aggression} = \sqrt{\frac{1}{t_1 - t_2} \int_{t_1}^{t_2} \left| \dot{\delta}(t) \right|^2 dt} \tag{3}$$

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

where:

- target is the target movement signal
- response is the response signal
- $\delta(t)$ is the pilot's input signal; $\dot{\delta}(t)$ is its time derivative
- t_1, t_2 are the start and end of a time interval for aggression analysis

The signals are actually available as discrete time series, with a sample rate of 60 Hz. To evaluate the performance and aggression in a certain time period (the whole task run or a specific portion, for example), their RMS is calculated:

$$\text{error}_{\text{RMS}} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}\left(\text{error}_i - \text{error}_{\text{mean}}\right)^2} \tag{4}$$

$$\text{aggression}_{\text{RMS}} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}\left(\text{aggression}_i - \text{aggression}_{\text{mean}}\right)^2} \tag{5}$$

Since experimental data were directly measured in terms of longitudinal and lateral components of the input, the composed total value of tracking error and input aggression can be calculated as follows:

$$\text{error}_{\text{total}} = \sqrt{\text{error}^2_{\text{longitudinal}} + \text{error}^2_{\text{lateral}}} \tag{6}$$

$$\text{aggression}_{\text{total}} = \sqrt{\text{aggression}^2_{\text{longitudinal}} + \text{aggression}^2_{\text{lateral}}} \tag{7}$$

For the evaluation of a complete task run, the RMS value is directly calculated, and the results are utilised as baseline data. For the evaluation of specific conditions within a task run, timestamps, when the conditions were met, are marked, and the corresponding error or aggression values are extracted to calculate the local RMS value. Data extracted from all task runs were grouped in terms of participants and task patterns respectively, to analyse the effect of differences among individuals and task patterns.

3.3.7 Results prediction

The randomness of the target movement and the presence of the boundary indicated that participants might find the tracking task at some times easy and difficult at others. The changing difficulties of the tasks might trigger participants to shift their input strategies, resulting in a change in input aggressiveness and tracking error. The possible results are listed below.

- **Tracking error**

  1. *The tracking error was lower in easy tasks than that in difficult tasks.* If this result should appear, it might indicate that the participant tended to focus on avoiding boundaries and treating tracking as a secondary task under critical boundary situations.
  2. *The tracking error was higher in easy tasks than that in difficult tasks.* This result might indicate that under difficult tasks, the participants force themselves to be more accurate in performing the tracking tasks, with the secondary effect of avoiding the boundary, resulting in a lower tracking error.

- **Input aggression**

  1. *The input aggression was lower in easy tasks than in difficult ones.* This might indicate that participants felt relaxed in easy tasks and did not need to make sudden adjustments for tracking

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

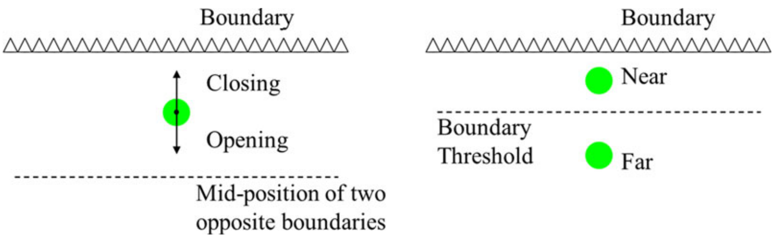

Figure 9. *Conditions of Section 4.1.*

tasks, while in difficult tasks participants felt the risk of hitting the boundary and adjusted their input strategy to complete the task.

2. *The input aggression was higher in easy tasks than that in difficult tasks.* This situation might happen when participants abandon tracking tasks and stay in a safe area during difficult tasks.

4.0 Data analysis

4.1 Conditions and boundary thresholds

Different *conditions* were defined to distinguish different groups of situations the participants encountered during the tasks:

(a) **Condition 1**

- *Closing*: the pointer is moving towards one of the boundaries
- *Opening*: the pointer is moving away from one of the boundaries

(b) **Condition 2**

- *Near*: the point is between one of the boundaries and its corresponding boundary threshold
- *Far*: the pointer is outside boundary thresholds

(c) **Condition 3**

- *Closing and Near*: both the *Closing* and *Near* conditions are simultaneously met
- *Opening or Far*: any of the *Opening* or *Far* conditions (or both) is met

They are illustrated in Fig. 9.

In this group of definitions, *boundary threshold* needs further clarification. The threshold should be small enough to only capture difficult task conditions, where participants are most likely to change their input strategy; however, it should be large enough to sample a sufficient amount of data for the analysis. For this reason, different boundary thresholds were tested. The duration of the *Closing and Near* condition in every task run, for a total of 216 test runs under different boundary thresholds, is plotted in Fig. 10. Each task run lasts 120 seconds.

When the boundary threshold is set to 0.1 (i.e., 10% of the cyclic scale), approximately 50% of the data is in the range of 27 s to 40 s, with a median value of 33 s. When it is set to 0.15, approximately 50% of the data is in the range of 40 s to 53 s, with a median value of 49 s. Subsequent analysis showed that using boundary threshold values of 0.1 or 0.15 had no significant influence on the results. The analysis presented in the following is based on a boundary threshold of 0.1.

4.2 Regression analysis

In this research, Gaussian process regression(GPR) [36] is used to analyse if the tracking error can be explained by experimental factors. Different task conditions were included in the regression model

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

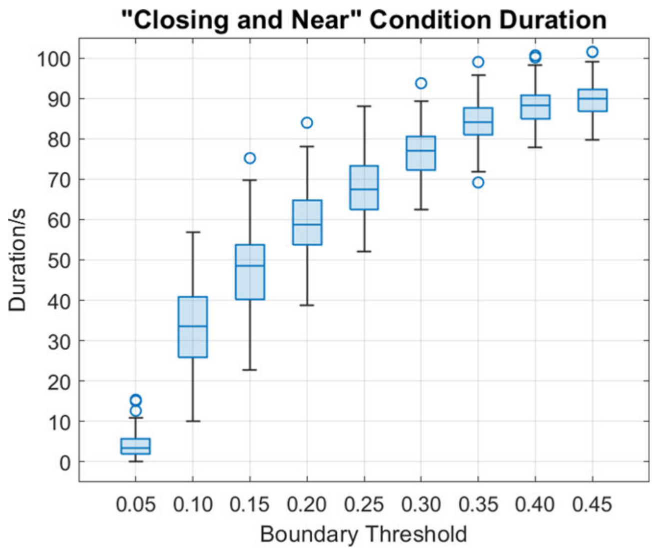

Figure 10. *Duration of Closing and Near condition for different values of boundary threshold.*

(participant's ID, task runs, task pattern, input aggression, task-failed flag) to investigate how these conditions affected the tracking errors.

The same basic functions and kernel functions were applied for longitudinal, lateral and composed total data, namely Equations (6) and (7). The scattered response plots, predicted vs actual plots, and residual plots were generated and presented to demonstrate the validity of the regression method. Then partial dependence plots were applied to interpret experimental factors.

4.3 Failing the task

The existence of critical boundaries meant that participants could fail the task before the defined task duration of 120 s was reached. Statistical and regression analysis was done with task runs that did not fail, to extract data that fully represented participants' performances under pressure. Task runs that failed will be treated individually to analyse how the test developed leading to the failure of the task.

5.0 Results

5.1 Individual analysis

5.1.1 Baseline performances

Baseline performances were extracted from the first three to five task runs where boundary avoidance tracking was not involved. The RMS values of the tracking error are shown in Fig. 11. The RMS values of the aggression are shown in Fig. 12. The ANOVA analysis results presented in Table 1 show that in terms of both the cyclic tracking error and input aggression, the participants presented significantly different behaviours ($p < 0.01$), which means the task design between point tracking and boundary avoidance tracking was effective in triggering different participants' control strategies to cope with the different task conditions.

In terms of tracking errors, the following results were observed from Fig. 11:

1. Among all the tracking tasks performed with the cyclic stick, the participants showed higher and more scattered tracking errors in tasks without BAT (M = 0.109, SD = 0.074, where M stands for mean and SD for standard deviation) than in those with BAT (M = 0.075, SD = 0.044). Input aggression was higher for task runs without BAT (M = 1.207, SD = 0.443), than for those with BAT (M = 1.137, SD = 0.520).

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

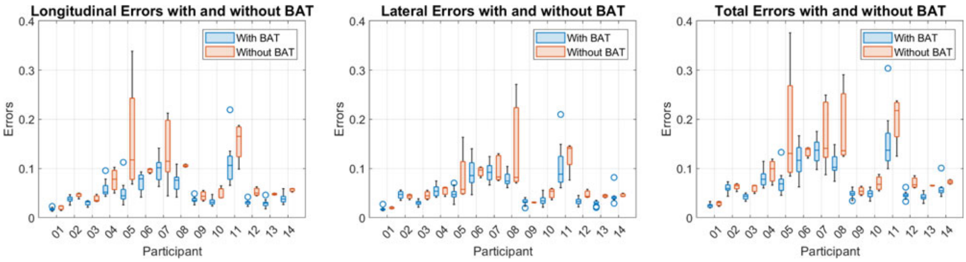

Figure 11. *RMS values of tracking error.*

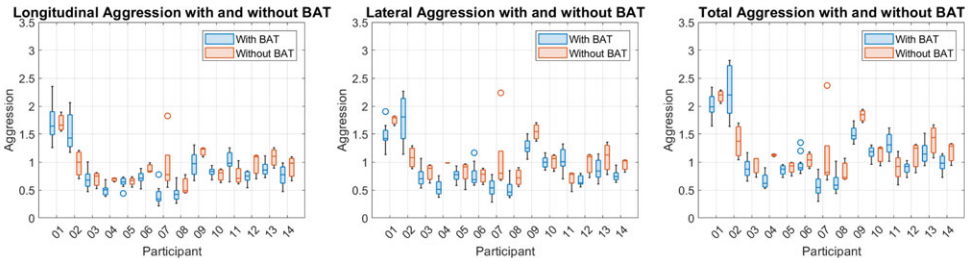

Figure 12. *RMS values of input aggression.*

2. In task runs without BAT, the tracking error of participants 4 and 7 was higher in the longitudinal than in the lateral direction; the tracking error of participant 8 was higher in the lateral than in the longitudinal direction; the tracking errors of all participants showed significant difference ($p < 0.01$) in the longitudinal direction, but not in the lateral direction ($p > 0.05$).
3. For most participants, task runs without BAT showed a higher and more scattered tracking error than that of task runs with BAT; exceptions are participants 1, 2, 6, 11, 13 and 14, whose tracking errors were less scattered in task runs without BAT, and participant 9, whose tracking error showed no significant difference.
4. Participant 11 is the only one whose tracking error in task runs with BAT was higher than that in task runs without BAT, but the largest error was treated as an outlier.

The comparison discussed above showed that a smaller tracking error occurred in BAT tasks. This was expected since tasks without BAT were meant for participants to get familiar with the joystick and the task. The absence of a fail condition made participants operate in a relaxed environment. During BAT tasks, instead, participants were given a workload higher than they expected, which on several occasions resulted in better performance, at least within certain ranges of parameters. The presence of the boundary also resulted in a less scattered tracking error in BAT tasks because participants were guided to track the target while avoiding the boundary.

The comparison of the aggression in Fig. 12 produced the following results:

1 Only participants 2 and 11 had a lower aggression in task runs without BAT than that in task run with BAT, and other participants had a higher aggression in task runs without BAT. The difference in the input aggression is significant ($p < 0.001$).

2 Scattered degree of aggression showed no significant difference for most cases except that participant 7 showed a larger scatter region (contributed by an outlier) in task runs with BAT. The changing of aggression for participants 2 and 11 was unique among other participants, indicating that they chose less aggressive input strategies for BAT tasks, while other participants utilised more cautious/careful input strategies for more difficult tasks.

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

Table 1. ANOVA analysis between with and without BAT tasks

Difference Significance: With and Without BAT Tasks		
Tracking Error		
Longitudinal		
Source	***F*-value**	***P*-value**
WithWithoutBAT	46.57	<0.001
Participant	24.67	<0.001
Interaction	3.28	<0.001
Lateral		
Source	***F*-value**	***P*-value**
WithWithoutBAT	13.85	<0.001
Participant	27.79	<0.001
Interaction	1.66	0.073
Total		
Source	***F*-value**	***P*-value**
WithWithoutBAT	37.22	<0.001
Participant	30.62	<0.001
Interaction	2.66	0.002
Input Aggression		
Longitudinal		
Source	***F*-value**	***P*-value**
WithWithoutBAT	5.24	0.023
Participant	50.75	<0.001
Interaction	6.05	<0.001
Lateral		
Source	***F*-value**	***P*-value**
WithWithoutBAT	10.38	0.002
Participant	42.01	<0.001
Interaction	7.20	<0.001
Total		
Source	***F*-value**	***P*-value**
WithWithoutBAT	4.20	0.042
Participant	65.79	<0.001
Interaction	8.49	<0.001

5.1.2 Analysis of conditions

(a) **Condition 1**
Figure 13 shows the tracking error and input aggression for *Closing* and *Opening* conditions. The tracking errors under the *Closing* condition (M = 0.069, SD = 0.035) are slightly larger than those under *Opening* (M = 0.045, SD = 0.029) for most participants, but the results showed did not show sufficient statistical significance ($p > 0.05$). Furthermore, participants 5, 6, 8, 11 and 14 presented a larger scattered region for tracking errors. Combined with

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

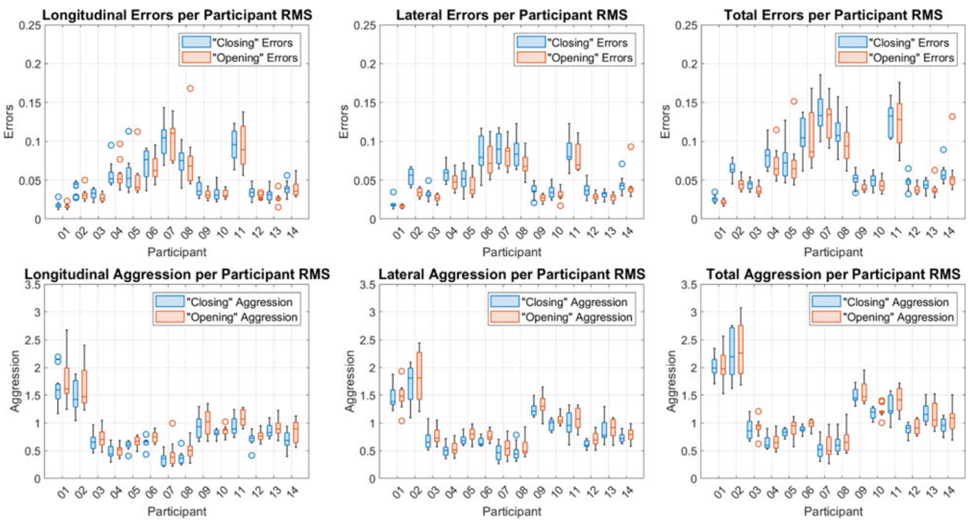

Figure 13. *Error and aggression under the closing and opening conditions.*

their behaviour during the test, a conclusion can be drawn that they tended to aggressively pull back the joystick when the target was leaving the boundary area, resulting in larger maximum tracking error values. The input aggression under the *Closing* condition was slightly larger than under *Opening*, though again no sufficient significance could be observed ($p > 0.05$).

The result shows that participants were more likely to control the stick less aggressively when they tried to follow the target getting close to the boundary, to avoid hitting it (Table 2).

(b) **Condition 2**

Figure 14 shows the tracking error and input aggression for *Far* and *Near* conditions. The tracking errors under the *Near* condition (M = 0.027, SD = 0.012) are significantly lower than under *Far* (M = 0.076, SD = 0.039) for all participants ($p < 0.001$). The analysis of the participants' behaviour during the test runs led to two possible conclusions:

1. The participants exerted greater effort to control the stick, achieving high accuracy for the point tracking task while preventing hitting the boundary.
2. When the target was in the threshold of the *Near* condition, it was stationary for a larger portion of time than outside the threshold, so point-tracking tasks were easier.

The input aggression showed inconsistent results across participants. For some of them, the input aggression was higher under the *Near* condition, while for others it was lower. Participants 10 and 11 presented different input strategies for the longitudinal and lateral direction. This might indicate that they tended to focus on one direction. For all participants, the difference in the input aggression was significant ($p < 0.05$) (Table 3).

(c) **Condition 3**

Figure 15 shows the tracking error and input aggression for *Closing and Near* and *Opening or Far* conditions. The tracking errors under the *Closing and Near* (M = 0.025, SD = 0.010) and *Opening or Far* (M = 0.072, SD = 0.038) conditions showed a similar trend as under *Near* and *Far* mainly because tracking errors under *Closing* and *Opening* showed no significant difference. The difference in tracking error here also showed statistical relevance ($p < 0.001$). Inconsistent results were observed for aggression also under this group. The difference between conditions and participants is significant for longitudinal and lateral axis ($p < 0.05$), but not the composed

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

Table 2. *ANOVA analysis between the Closing and Opening conditions*

Difference Significance: Closing and Opening conditions		
Tracking Error		
Longitudinal		
Source	***F*-value**	***P*-value**
Condition	4.15	0.042
Participant	58.58	<0.001
Interaction	0.29	0.993
Lateral		
Source	***F*-value**	***P*-value**
Condition	33.78	<0.001
Participant	70.93	<0.001
Interaction	1.03	0.421
Total		
Source	***F*-value**	***P*-value**
Condition	13.80	<0.001
Participant	76.19	<0.001
Interaction	0.45	0.951
Input Aggression		
Longitudinal		
Source	***F*-value**	***P*-value**
Condition	18.32	<0.001
Participant	89.52	<0.001
Interaction	0.42	0.963
Lateral		
Source	***F*-value**	***P*-value**
Condition	14.38	<0.001
Participant	98.80	<0.001
Interaction	0.17	1.000
Total		
Source	***F*-value**	***P*-value**
Condition	4.97	0.027
Participant	132.63	<0.001
Interaction	0.24	0.997

total value ($p > 0.05$). Different participants applied different input strategies under severe task conditions, resulting in different point-tracking performances (Table 4).

5.2 Task pattern analysis

Figure 16 shows plotted each task run and their task pattern. Table 5 lists the mean and standard deviation of each task pattern. As presented in Table 6, tracking errors had no significant difference for each participant across different task patterns ($p > 0.05$), while aggression showed a great difference ($p < 0.001$),

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

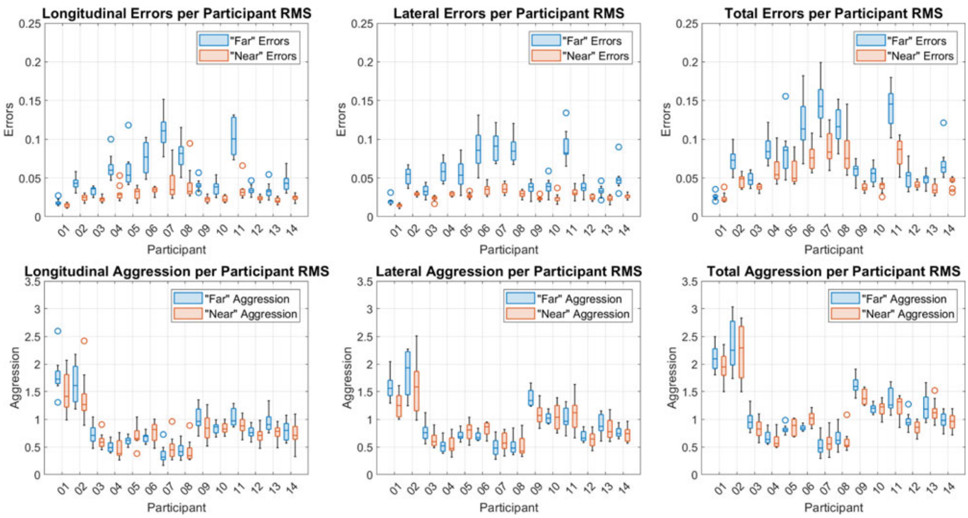

Figure 14. *Error and aggression under the Near and Far conditions.*

which means that participants tried to adjust their input strategy to cope with different task patterns, resulting in similar tracking errors (Fig. 17).

According to the definition and design of each task pattern, where Task 3 and 4 added a shifting movement for both target and boundary, the difficulty of Tasks 3 and 4 is higher than that of Tasks 1 and 2, but the performance varies in terms of participants. For example, participants 2 and 12 showed no significant difference in performing different task patterns; participants 5, 6, 7, 13 and 14 presented the largest tracking error when the BAT task was first introduced; this could mean that some learning effect still had an impact on the participants' performance during several of the first BAT tasks; participants 4, 9, 10 presented the largest tracking error when the shifting movement of the target and boundary was first introduced. For each task pattern, participants always showed some decrease in tracking error with the prosecution of the task runs, which indicates that some learning effect was present throughout the whole experiment.

The input aggression also showed variations. Participants 2, 3, 8, 9 presented the largest input aggression in Task 4, which should be the most difficult by design. Other participants showed a broader variety of results. Specifically, participant 7 showed the largest input aggression in Task 1, which was the first group of tests where the BAT task was introduced. This might be explained by the familiarisation with the BAT task as the experiment proceeded, which is consistent with the decreasing tracking error after Task 1. Still, participant 7 was the one with the largest tracking error.

5.3 Failed task runs analysis

In this section, failed task runs that were representative are plotted and analysed.

In the test run shown in Fig. 18, the participant failed the task at about 65 s. Between 42 s and 50 s, the lateral direction target was approaching the right boundary. While the participant tried to track the target, they pulled the stick back periodically to avoid exceeding the boundary. As a consequence, the pointer also approached the boundary periodically, but with some delay with respect to the participant's input. Right before the pointer hit the boundary, while approaching it, the participant tried to slow down or stop the pointer by pulling the stick backwards. Because of the lag caused by the transfer function, the response was delayed, resulting in hitting the target, and thus failing the task.

Another failed task run is shown in Fig. 19. The task failed at around 17 s. Comparing the input signal in both longitudinal and lateral directions, one can observe that the participant was trying to control the stick to follow the target in the lateral direction, where the target underwent a sudden stop close to

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

Table 3. *ANOVA analysis between the Near and Far conditions*

Difference Significance: Near and Far Conditions		
Tracking Error		
Longitudinal		
Source	***F*-value**	***P*-value**
Condition	335.44	<0.001
Participant	43.93	<0.001
Interaction	13.48	<0.001
Lateral		
Source	***F*-value**	***P*-value**
Condition	519.03	<0.001
Participant	44.27	<0.001
Interaction	19.71	<0.001
Total		
Source	***F*-value**	***P*-value**
Condition	162.14	<0.001
Participant	63.15	<0.001
Interaction	5.24	<0.001
Input Aggression		
Longitudinal		
Source	***F*-value**	***P*-value**
Condition	12.34	<0.001
Participant	70.31	<0.001
Interaction	2.10	0.014
Lateral		
Source	***F*-value**	***P*-value**
Condition	10.77	0.001
Participant	72.41	<0.001
Interaction	3.04	<0.001
Total		
Source	***F*-value**	***P*-value**
Condition	9.26	0.003
Participant	142.73	<0.001
Interaction	1.10	0.355

the boundary. In the meanwhile, in the longitudinal direction, the target was also moving towards the boundary, but the participant seemed to ignore this axis and focused their attention on the lateral axis, consequently failing the task with respect to the longitudinal axis.

5.4 Regression analysis

During data processing, it was observed that participants who exhibited higher levels of input aggression presented lower tracking errors. Conversely, participants who had higher tracking error and task failure

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

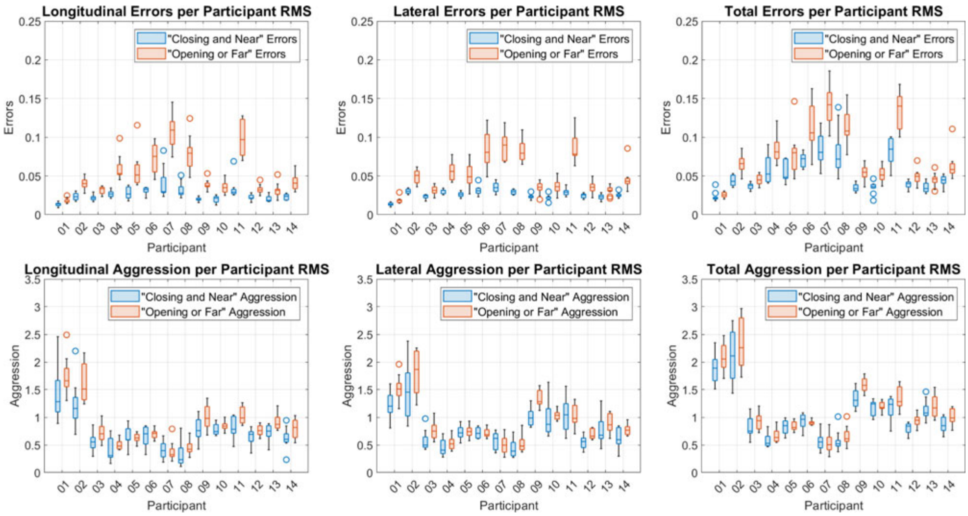

Figure 15. *Error and aggression under the Closing and Near and Opening or Far conditions.*

count exhibited lower input aggression. In order to provide a more comprehensive analysis of the data, all data points were plotted in Fig. 20, including failed trials. The figure indicates that although there is considerable variation in the data points, they tend to be concentrated in the lower-left corner of the graph. This pattern suggests that tracking error and input aggression may negatively correlate. To further investigate this relationship, GPR was utilised.

MATLAB 2022b was utilised to use its built-in functions to script an Optimisable GPR code. During the analysing procedure, hyper-parameters for GPR were optimised using Bayesian optimise method. Figures 21–23 show the results of the GPR. Table 7 shows the root mean square error (RMSE), R-squared, mean squared error (MSE), and mean average error (MAE) of the prediction.

Partial dependence represents the relationships between predictor variables and predicted responses in a trained regression model. It computes the partial dependence of predicted responses on a subset of predictor variables by marginalising over the other variables [37]. In this paper, partial dependence was used to explain the effect of predictors (participant's ID, task runs, task pattern, input aggression, task-failed flag) on the predicted response (tracking error).

The participant number (1 to 14), task runs (1 to 17), task pattern (0 to 4), task failure (0 or 1), and input aggression were used as the predicting variables while tracking error was used as the response. Regression figures are plotted in Figs 21–23. Training results are shown in Table 7. According to the plots and results table, the current model variables can explain about 50%–60% of the variance of the tracking error. The variables are plotted separately to investigate the partial dependence of the tracking error.

Figures 24–28 show each variable in relation to the tracking error for longitudinal, lateral axes and total error. The Participants partial factor showed the most variance across all the variables (a range from 0.02 to 0.12, a margin of about 0.1 of the tracking error), which could be understood because each individual had different proficiency with the stick, as well as a different understanding of the task. The task runs showed that although all tasks were random and different from each other, there was still a slight negative correlation with the tracking error (a margin of about 0.06), which seems to indicate that a learning effect still existed across the trials. This could be explained by familiarisation with the joystick, and some remaining implicit similarities between the tasks. Task patterns seem to have a weak positive relation with the tracking error (a margin of about 0.02), meaning that difficulty increased across the five task patterns. Task failure increases the tracking error by about 0.02.

Table 4. *ANOVA analysis between Closing and Opening and Near or Far conditions*

Difference Significance: Closing and Near and Opening or Far Conditions		
Tracking Error		
Longitudinal		
Source	***F*-value**	***P*-value**
Condition	392.54	<0.001
Participant	45.36	<0.001
Interaction	15.91	<0.001
Lateral		
Source	***F*-value**	***P*-value**
Condition	537.74	<0.001
Participant	50.57	<0.001
Interaction	23.29	<0.001
Total		
Source	***F*-value**	***P*-value**
Condition	167.23	<0.001
Participant	68.64	<0.001
Interaction	5.79	<0.001
Input Aggression		
Longitudinal		
Source	***F*-value**	***P*-value**
Condition	42.93	<0.001
Participant	62.66	<0.001
Interaction	1.95	0.025
Lateral		
Source	***F*-value**	***P*-value**
Condition	35.88	<0.001
Participant	66.40	<0.001
Interaction	2.11	0.014
Total		
Source	***F*-value**	***P*-value**
Condition	25.72	<0.001
Participant	140.93	<0.001
Interaction	1.01	0.444

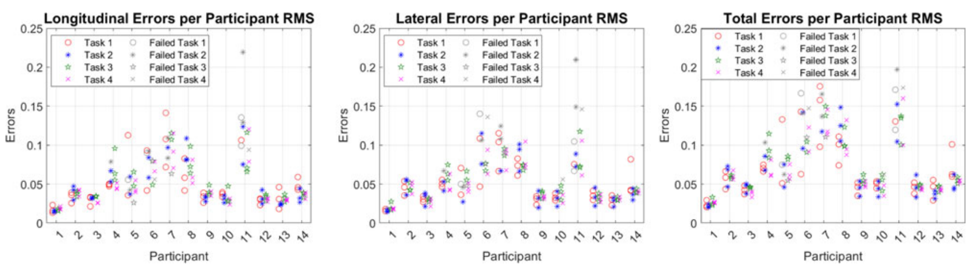

Figure 16. *Tracking error under different tasks.*

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

Table 5. *Mean and standard deviation of tracking error and input aggression among different task patterns*

	Task 1		Task 2		Task 3		Task 4	
	Error	Aggr.	Error	Aggr.	Error	Aggr.	Error	Aggr.
Mean	0.074	1.104	0.080	1.117	0.073	1.141	0.071	1.187
Standard Deviation	0.043	0.452	0.057	0.460	0.034	0.560	0.039	0.606

Table 6. *ANOVA analysis among different task patterns*

Difference Significance: Task Patterns		
Tracking Error		
Longitudinal		
Source	***F*-value**	***P*-value**
Task Pattern	1.49	0.220
Participant	32.93	<0.001
Interaction	0.89	0.661
Lateral		
Source	***F*-value**	***P*-value**
Task Pattern	0.39	0.759
Participant	28.62	<0.001
Interaction	0.71	0.893
Total		
Source	***F*-value**	***P*-value**
Task Pattern	0.78	0.508
Participant	38.23	<0.001
Interaction	0.78	0.804
Input Aggression		
Longitudinal		
Source	***F*-value**	***P*-value**
Task Pattern	8.15	<0.001
Participant	102.66	<0.001
Interaction	3.65	<0.001
Lateral		
Source	***F*-value**	***P*-value**
Task Pattern	0.67	0.574
Participant	82.24	<0.001
Interaction	3.05	<0.001
Total		
Source	***F*-value**	***P*-value**
Task Pattern	3.04	0.032
Participant	156.85	<0.001
Interaction	4.25	<0.001

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

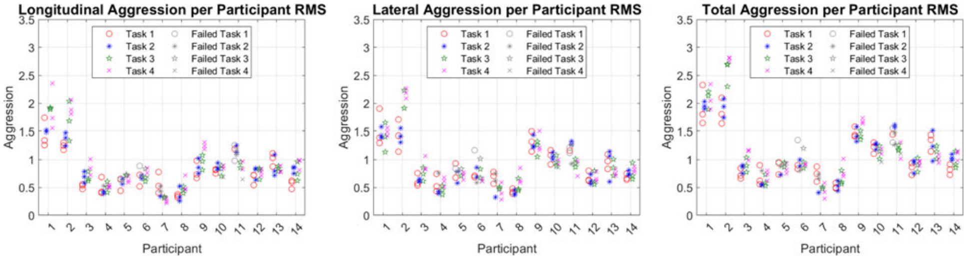

Figure 17. *Input aggression under different tasks.*

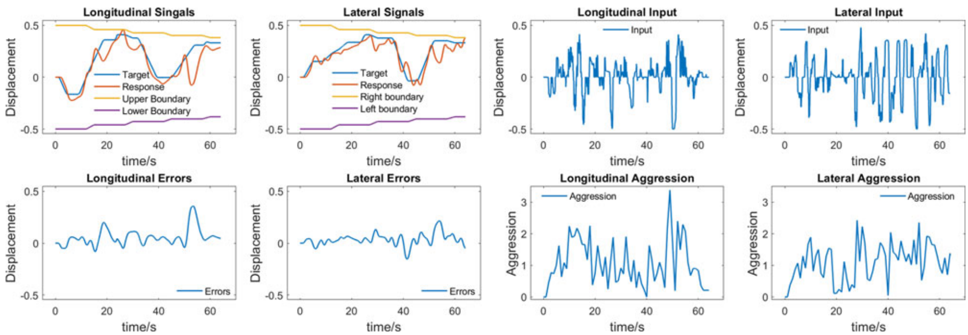

Figure 18. *Task failed while trying to track the target near the boundary.*

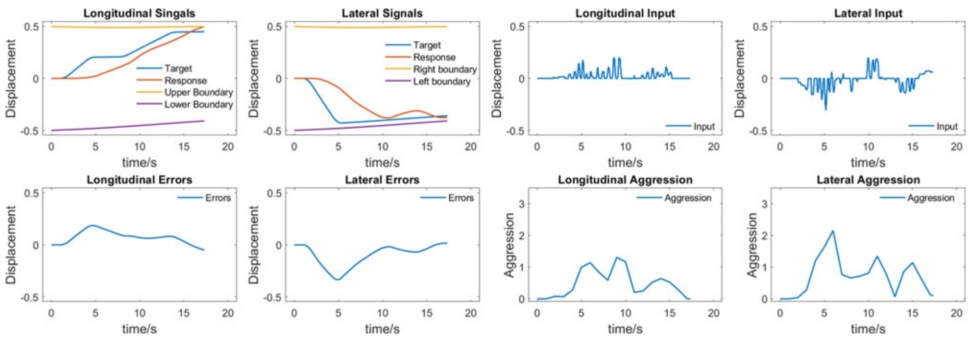

Figure 19. *Task failed while ignoring one axis.*

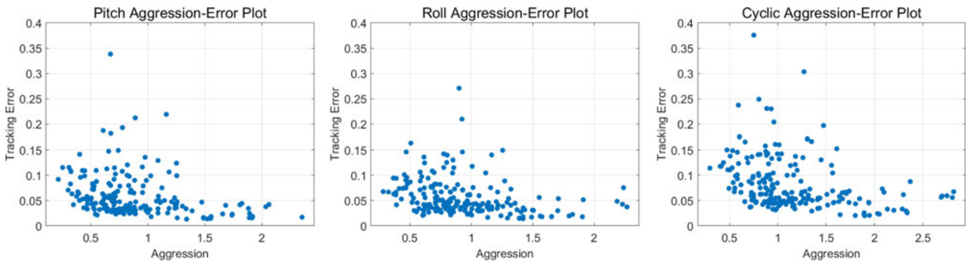

Figure 20. *Aggression-error plot.*

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

Table 7. *Regression training results*

	Longitudinal	Lateral	Total
RMSE	0.027	0.025	0.033
R-squared	0.59	0.53	0.62
MSE	7.2e-4	6.0e-4	1.1e-3
MAE	0.015	0.014	0.019

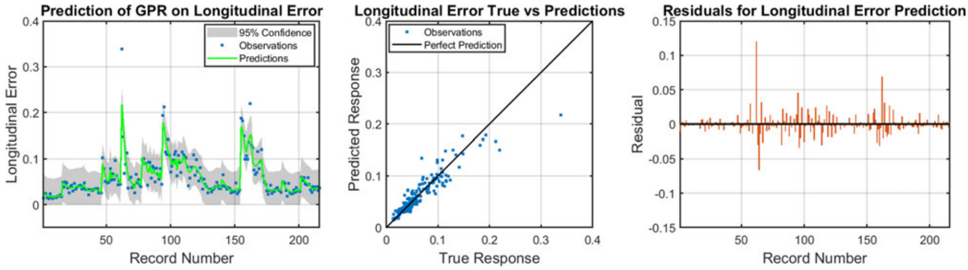

Figure 21. *Gaussian process regression in the longitudinal axis.*

Figure 22. *Gaussian process regression in the lateral axis.*

Figure 23. *Gaussian process regression for total data.*

Input aggression showed an inconsistent correlation with the tracking error. For the longitudinal axis, the increase in aggression also increased the tracking error, which ranged at about 0.006, while increasing the input aggression at the lateral axis reduced the tracking error with a margin of about 0.04. The combined total error showed a negative correlation between input aggression and tracking error, with a margin of about 0.06.

The tracking error was affected by multiple factors. The above-mentioned analysis may have not covered all aspects of the problem, but it shows some interesting results. The inconsistent correlation for the aggression deserves further investigation in future research.

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

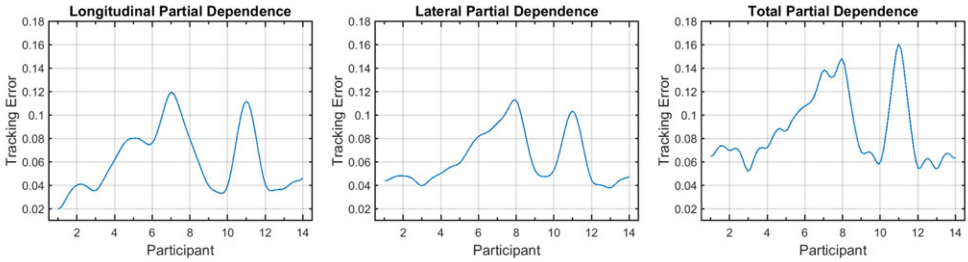

Figure 24. *Partial dependence of participants.*

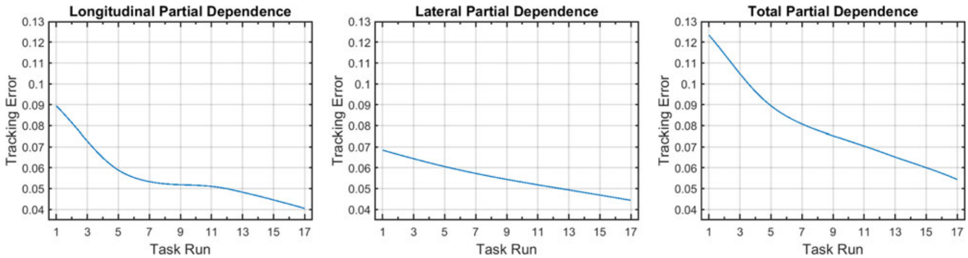

Figure 25. *Partial dependence of task runs.*

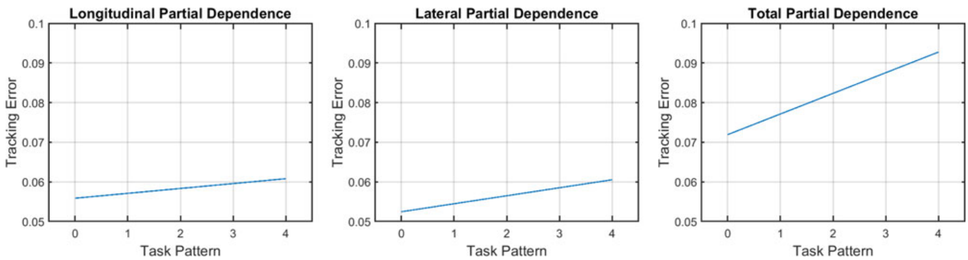

Figure 26. *Partial dependence of task patterns.*

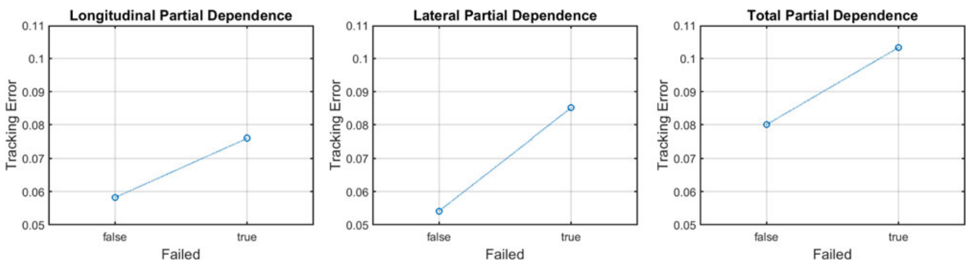

Figure 27. *Partial dependence of failure.*

6.0 Discussion

This study aimed at investigating tracking error and input aggression under different task conditions in point tracking and boundary avoidance tasks. A regression model was implemented to find possible relationships between input aggression and tracking error. The following aspects are discussed according to the questions raised in this paper.

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

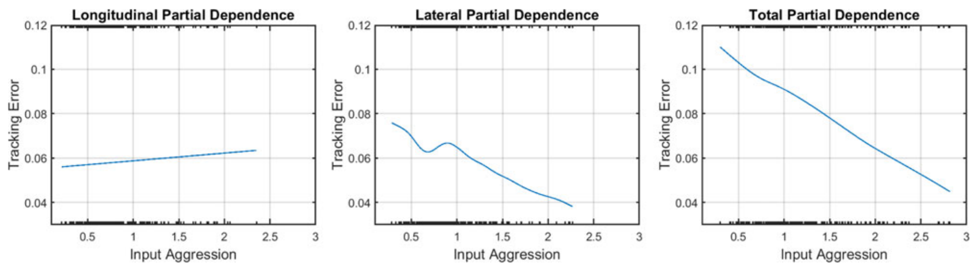

Figure 28. *Partial dependence of aggression.*

6.1 Participants and task patterns

The volunteers that participated in the experiment had no previous experience in either helicopter piloting or similar simulation tasks.

The participants could familiarise themselves with the task configuration and joystick manipulation through Task 0, which they could try as many times as they wished. Therefore, it could be assumed that participants had sufficient experience and confidence to deal with more challenging tasks before entering the BAT task. However, different participants still exhibited different input strategies and tracking performance, which were related to both the task configuration and the participants themselves.

From the perspective of task configuration:

1. In this study, the simulated missions were composed of a critical and a sub-critical task. Considering its impact on the results, BAT should be considered the critical task, as its failure would directly lead to the termination of the mission, whereas point-tracking can be considered sub-critical because even if the participants could not keep the point-tracking cursor in the desired or acceptable regions, the mission would not fail. This task configuration might affect the participants' manipulation decisions, which answers RQ1 mentioned in section 2.0.
2. The termination condition (i.e., participants' contact with the boundary) of the simulated task itself would also affect their strategies. We can assume that if contacting the boundary does not immediately terminate the task but affects the score after the task ends, participants will adopt different input strategies.
3. The distance and direction of the point-tracking target from the boundary significantly affected participants' manipulation strategies, regardless of different participants, which was also expected in the experiment design phase of this study.

In terms of participants, different task conditions resulted in different task performance, thus answering RQ2. The reason of this difference could be:

- Different participants had biased understandings of the task because it is highly abstract, without a standardised code of conduct. The participants' understanding and execution of the task were entirely determined by their personal cognitive capabilities.
- In Task 0, the participants might have not fully mastered the joystick manipulation, possibly owing to their inability to observe the difficulty of the subsequent tasks. Consequently, they could have been unable to judge the adequacy of their current proficiency for the subsequent simulated tasks, resulting in the need for further training, and thus a potential learning effect, during the whole experiment (the learning effect would have been limited to joystick proficiency rather than task configuration as in [12, 25], where periodic targets were adopted, allowing participants to anticipate the movement of the targets).
- Although all participants encountered the same task configuration sequence, their input strategies for the same task condition were different, possibly leading to different point-tracking task performances, thus answering RQ3.

https://doi.org/10.1017/aer.2024.35 Published online by Cambridge University Press

6.2 Limitations

A rather simple experimental setup was used in this study. This is expected to be insufficient to give the participants the possibility to feel the realism and immersion realistic, albeit simulated, helicopter tasks should have, possibly resulting in a too-relaxed attitude towards the experiment. In the study of [26], an appropriate/medium workload could improve the task performance of the participants. We cannot assume that in this study all participants immersed themselves in the simulated flight task to the same extent. However, even if more realistic simulation equipment were used, the differences in the participants' immersion could not be completely eliminated. However, improving the simulation environment should reduce the differences by an appreciable amount.

From the perspective of simulated task design, a simplified simulated task may be able to meet the research aims and make the data collection easy. However, whether such a task can accurately reflect the characteristics of a real one and reproduce the situations that may be encountered in real-world cases needs further verification. In addition, the task design itself has space for improvement. Despite the randomness of the target movement, one can observe from Fig. 8 that it became much more predictable near the boundary. Sudden movements were only presented for large distances between the target and the boundaries, possibly contributing to the lower tracking error under the *Near* condition. Improved tasks are being design tested in an ongoing study.

The main indicators used in this study to analyse the participants' performance are input aggression and tracking error, and the interpretation of the data mainly relies on statistical methods. The validity and specificity of the p-value are questioned [38, 39]. However, the p-value threshold for hypothesis testing has been the standard in the field and is generally reported across the literature spanning some 30 years [18]. The authors believe that this method and these indicators are effective within the scope of the present study. More effective methods will be tried in future research. In addition, more data from various aspects of the experiments will give the opportunity to address more faces of the problem.

There is still room for interpretation in the model fitting of input aggression and tracking error. The predictive function of Gaussian process regression should be able to provide a reference for the participants' input strategy.

Despite the aforementioned limitations, this study provides valuable preliminary task design and methods that will be exploited in future research on pilot involuntary response and the dependence of BDFT on task complexity.

7.0 Conclusions

This paper featured a simulation task design based on the concepts of point tracking and boundary avoidance tracking, and a data analysis method to investigate pilots' point tracking performance and input aggression. Several results could be drawn from this study. First, during the whole experiment, participants had a learning curve for the joystick and task pattern, reflected in a significantly reduced tracking error from the task without boundary to tasks with boundary. Then, after the boundary avoidance tracking was introduced in the task, participants presented different input strategies, detected through aggression, resulting in different levels of tracking error. When the target was near the boundary, the participants presented significantly lower tracking errors ($p < 0.001$), and the levels of input aggression were also lower for the longitudinal and lateral axis ($p < 0.05$). While the tracking error was not significantly affected by task patterns ($p > 0.05$), the input aggression showed a significant difference ($p < 0.001$). During the experiment, task failures were sometimes caused by two types of situations: (a) the delay between the manual input signal and the motion of the pointer, and (b) the distribution of the participants' attention. Regression analysis showed that tracking errors can be predicted to a certain extent, giving useful indications for pilot training and input strategy advising in future experiments.

This research suggests the existence of a relationship between task condition, input strategy and task performance. The random task design will be developed further to address the aforementioned limitations and serve as a template for future research. Input aggression during task performance could be an indicator for investigating BDFT and the trigger of PIO phenomena. The results of this research could serve as a database for the development of a pilot model, which describes pilots' behaviour and response to BAT tasks, and furthermore, the research on BDFT and rigid-body and aeroelastic APCs and RPCs. Future work will focus on the following aspects: implementing a more realistic flight simulation environment, with real helicopter control inceptors, heads-up display, and dashboards; utilising realistic helicopter dynamics and intuitive graphical flight task presentation; establishing a pilot model for pilot behaviour investigation and human-machine interaction modelling; investigating how control strategies affect BDFT and the occurrence of PIOs; and exploring the full potential of the measured data.

Competing interests. The first author, Qiuyang Xia, acknowledges the support from the China Scholarship Council (grant number: 202106020003). The third author, Pierangelo Masarati, the co-supervisor of Qiuyang Xia and the supervisor of Davide Marchesoli, is a member of the Journal's Editorial Board. The fourth author, Meng Liu, is the co-supervisor of Qiuyang Xia.

References

[1] Pavel M.D., Jump M., Dang-Vu B., et al. Adverse rotorcraft pilot couplings—past, present and future challenges. *Prog. Aerosp. Sci.*, 2013, **62**, pp 1–51.
[2] McRuer D.T. *Understanding and Preventing Unfavourable Pilot-Vehicle Interactions*. Washington, DC, USA: National Academy Press, 1997.
[3] Padfield G.D. and Lu L. The potential impact of adverse aircraft-pilot couplings on the safety of tilt-rotor operations. *Aeronaut. J.*, 2022, **126**, (1304), pp 1617–1647.
[4] Military Specification. Flying qualities of piloted airplanes. Technical report, MIL-F-8785C, 1980.
[5] Military Standard. Flying qualities of piloted aircraft. Technical report, US Dept. of Defense MIL-STD-1797A, 1990.
[6] Gray W. Boundary-escape tracking: A new conception of hazardous PIO. Technical report, Air Force Flight Test Center Edwards AFB, CA, USA, 2004.
[7] Gray W. Boundary avoidance tracking: a new pilot tracking model. In *AIAA Atmospheric Flight Mechanics Conference and Exhibit*, San Francisco, CA, USA, AIAA, 2005, p 5810.
[8] Wilson J.C., Nair S., Scielzo S. and Larson E.C. Objective measures of cognitive load using deep multi-modal learning: a use-case in aviation. *Proc. ACM Interact. Mobile Wearable Ubiquit. Technol.*, 2021, **5**, (1), pp 1–35.
[9] Babu M.D. Investigation of boundary avoidance tracking theory from ocular parameters through simulator and inflight studies. *l'Aerotecnica, Missili Spazio* 2021, **100**, (4), pp 303–319.
[10] Mitchell D.G. and Klyde D.H. This is pilot gain. In *AIAA Scitech 2019 Forum*, San Diego, CA, USA, AIAA, 2019, p 0562.
[11] Lu L. and Jump M. Pilot modelling for boundary hazard perception and reaction study. In *43rd European Rotorcraft Forum, ERF 2017*, Milano, Italy, 2017, 1, pp 640–652.
[12] Ji H., Lu L., White M.D. and Chen R. Advanced pilot modeling for prediction of rotorcraft handling qualities in turbulent wind. *Aerosp. Sci. Technol.*, 2022, **123**, p 107501.
[13] Gratton G.B. and Bromfield M.A. Aspects of General Aviation flight safety research. In *RAeS GA Conference 2008; Gratton and Bromfield on Safety Research*, Cambridge, Royal Aeronautical Society, 2008, pp 1–12.
[14] Craun R., Acosta D.M., Beard S.D., Hardy G., Leonard M.W. and Weinstein M.J. Boundary avoidance tracking for instigating pilot induced oscillations. In *AIAA Atmospheric Flight Mechanics (AFM) Conference*, Boston, MA, USA, AIAA, 2013, p 4688.
[15] Afloare A.-I. Pilot induced oscillations detection using boundary avoidance tracking procedure. *Sci. Bull. Ser. D*, 2015, **77**, (4), pp 27–34.
[16] Wilson J.C. *Cognition and Context-Aware Computing: Towards a Situation-Aware System with a Case Study in Aviation*. PhD thesis, Southern Methodist University, 2020.
[17] Endsley M.R. Toward a theory of situation awareness in dynamic systems. *Hum. Fact.: J. Hum. Fact. Ergonom. Soc.*, 1995, **37**, (1), pp 32–64.
[18] Endsley M.R. A systematic review and meta-analysis of direct objective measures of situation awareness: a comparison of SAGAT and SPAM. *Hum. Fact.*, 2021, **63**, (1), pp 124–150.
[19] Endsley M.R. *Sagat: A Methodology for the Measurement of Situation Awareness (Nor Doc 87–83)*. Hawthorne, CA: Northrop Corporation, p 18, 1987.
[20] Durso F.T., Hackworth C.A., Truitt T.R, Crutchfield J., Danko N. and Manning C.A. Situation awareness as a predictor of performance for en route air traffic controllers. *Air Traffic Control Q.*, 1998, **6**, (1), pp 1–20.
[21] de Winter J.C.F., Eisma Y.B., Cabrall C.D.D., Hancock P.A. and Stanton N.A. Situation awareness based on eye movements in relation to the task environment. *Cognit. Technol. Work*, 2019, **21**, (1), pp 99–111.
[22] Bolstad C.A, Foltz P., Franzke M., Cuevas H.M., Rosenstein M. and Costello A.M. Predicting situation awareness from team communications. In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*, Sage, 2007, pp 789–793.
[23] Jones D.G. and Endsley M.R. Can real-time probes provide a valid measure of situation awareness. In *Proceedings of the Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium*, Savannah, GA, USA, 2000, pp 1–6.

[24] Farley T.C., Hansman R.J., Amonlirdviman K. and Endsley M.R. Shared information between pilots and controllers in tactical air traffic control. *J. Guid. Control Dyn.*, 2000, **23**, (5), pp 826–836.
[25] Lu L. and Jump M. Multiloop pilot model for boundary-triggered pilot-induced oscillation investigations. *J. Guid. Control Dyn.*, 2014, **37**, (6), pp 1863–1879.
[26] Feng C., Wanyan X., Yang K., Zhuang D. and Wu X. A comprehensive prediction and evaluation method of pilot workload. *Technol. Health Care*, 2018, **26**, (S1), pp 65–78.
[27] Feng C., Wanyan X., Liu S., Chen H. and Zhuang D. An approach to situation awareness (SA) assessment in flight simulation: SA dynamic circulation (SADC) model. *Hum. Fact. Ergonom. Manuf. Serv. Ind.*, 2021, **31**, (5), pp 559–569.
[28] Feng C., Liu S., Wanyan X., Chen H., Min Y. and Ma Y. EEG feature analysis related to situation awareness assessment and discrimination. *Aerospace*, 2022, **9**, (10), p 546.
[29] Zanoni A., Cocco A. and Masarati P. Multibody dynamics analysis of the human upper body for rotorcraft– pilot interaction. *Nonlinear Dyn.*, 2020, **102**, (3), pp 1517–1539.
[30] Daniele M., Quaranta G., Masarati P. and Zanoni A. Pilot in the loop simulation of helicopter-ship operations using virtual reality. *l'Aerotecnica, Missili Spazio* 2020, **99**, (1), pp 53–62.
[31] Zanoni A., Zago M., Paolini R., Quaranta G., Galli M. and Masarati P. On task dependence of helicopter pilot biodynamic feedthrough and neuromuscular admittance: an experimental and numerical study. *IEEE Trans. Hum.-Mach. Syst.*, 2021, **51**, (5), pp 421–431.
[32] Zanoni A., Cocco A., Marchesoli D., Talamo C., Masarati P., Colombo F., Kemp S. and Fosco E. Pilot biomechanics for the definition of a rotorcraft-pilot interaction experiment. In *78th International Annual Forum Vertical Flight Society*, The Vertical Flight Society, 2022, pp 1–8.
[33] Muscarello V., Quaranta G. and Masarati P. The role of rotor coning in helicopter proneness to collective bounce. *Aerosp. Sci. Technol.*, 2014, **36**, pp 103–113.
[34] Jones M., Jump M., Lu L., Yilmaz D. and Pavel M.D. Using the phase-aggression criterion to identify rotorcraft pilot coupling events. In *38th European Rotorcraft Forum*, Amsterdam, The Netherlands, September 4–7, 2012, pp 1–16.
[35] Gray W. A generalized handling qualities flight test technique utilizing boundary avoidance tracking. In *2008 U.S. Air Force T&E Days*, Los Angeles, CA, USA, AIAA, 2008, pp 1–19.
[36] Rasmussen C.E. and Williams C.K.I. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
[37] Friedman J.H. Greedy function approximation: a gradient boosting machine. *Ann. Statist.*, 2001, **29**, (5), pp 1189–1232.
[38] Kline R.B. *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*. Washington, DC: American Psychological Association, 2005.
[39] Cumming G. The new statistics: why and how. *Psychol. Sci.*, 2014, **25**, (1), pp 7–29.

Cite this article: Xia Q., Marchesoli D., Masarati P. and Liu M. (2024). Pilot performance during simulated point and boundary avoidance tracking tasks. *The Aeronautical Journal*, **128**, 2191–2217. https://doi.org/10.1017/aer.2024.35