E.02.24-MOTA-D2.1 Functional HMI0: Experiment 1 Baseline Results

Document information	
Project Title	Project Modern Taxiing
Project Number	E.02.24
Project Manager	ENAC
Deliverable Name	Functional HMI0: Experiment 1 Baseline Results
Deliverable ID	D2.1
Edition	1.0
Template Version	03.00.00
Task contributors	
ISAE, ENAC	

Abstract

This document describes the development and execution of the first of three experiments for the Project Modern Taxiing (MoTa) validation campaign, which were run using the first version of MoTa interface. Two 35-minute scenarios, Medium and Hard, each featuring different operational events such as pilot error, restricted zone, towed aircraft, closed taxiway, and change in configuration, were performed with 12 air traffic ground controllers from around Europe. The results of this baseline experiment will be compared to the results of experiments 2 and 3, which will note the effect of a tactile interface and the use of taxibots, respectively. Current results indicate that the two scenarios are significantly different with respect to taskload, thus enabling one to determine the specific effect of the MoTa platform across a wide range of workload.

Authoring & Approval

Prepared By - *Authors of the document.*		
Name & Company	Position & Title	Date
Zarrin Chua, ISAE-Supaero	Post Doctorate	15/04/2015

Reviewed By - *Reviewers internal to the project.*		
Name & Company	Position & Title	Date
Mickael Causse, ISAE-Supaero	Assistant Professor	10/07/2015
Mathieu Cousy, ENAC	Project Leader	20/07/2015

Reviewed By - *Other SESAR projects, Airspace Users, staff association, military, Industrial Support, other organisations.*		
Name & Company	Position & Title	Date

Approved for submission to the SJU By - *Representatives of the company involved in the project.*		
Name & Company	Position & Title	Date

Rejected By - *Representatives of the company involved in the project.*		
Name & Company	Position & Title	Date

Rational for rejection
None.

Document History

Edition	Date	Status	Author	Justification
V 0.1	15/04/2015	Draft	Z.Chua	New Document
V 0.2	16/07/2015	Draft	Z.Chua	Modifications after project members reviews
V 1.0	20/07/2015	Release	M.Cousy	Modifications after project members reviews

Table of Contents

List of tables

List of figures

1 Executive summary

The baseline experiment, in addition to understanding the shortcomings of today's air traffic control technology in France, was also used to validate the two scenarios, Medium and Hard. The experiment was conducted at the Ecole Nationale de l'Aviation Civile (Toulouse, France) with 12 air traffic controller instructors with experience from airports around France. Five participants were from Charles-de-Gaulle, the airport that the simulation is based on. Three were from Paris Orly Airport, 1 from Melun Villaroche Aerodrome, 1 from Goteborg Landvetter Airport, 1 from Strasbourg International Airport, and 1 from Toulouse–Blagnac Airport.

Each participant performed a 30 minute practice scenario and two 35-minute ground taxiing scenarios, Medium and Hard. The run order was varied for each participant. Both scenarios varied by the number of aircraft and different operational events (restricted zone, pilot errors, closed taxiway, change in configuration, towed aircraft). The Medium aircraft featured a maximum of 31 aircraft whereas the Hard scenario had 51 aircraft. The Hard scenario featured all five operational events; the Medium had all except for the change in configuration. The change in configuration was planned to occur 15 minutes into the scenario, with a warning at 5 minutes, from the West configuration to the East. The Medium scenario was conducted strictly in the West configuration. During the experiment, the participants were outfitted with an electroencephalogram, an electrocardiogram, and an eyetracker. After each run, participants were asked to report their workload and situation awareness using the NASA Task Load Index and the Situation Awareness Rating Technique. They were also asked about their trust in the automation of their home airport.

The simulator was validated using a post-hoc 5-point Likert scale, with participants rating the simulator favourably (3.46 out of 5, 5 being a perfect simulator). Participants were also asked to compare against other simulators, notably, the one used at Charles-de-Gaule (4.75/5). The simulator execution was also validated, with the pseudopilot variable having no significance and the run order having mild significance on only the self-reported workload.

The results of this experiment demonstrate that the two scenarios are significantly different with respect to scenario complexity. Participants correctly treated a smaller percentage of aircraft in the Hard scenario compared to the Medium scenario. They reported having more workload and less situation awareness in Hard than Medium. The average taxiing experience (i.e. normalized deviation from the ideal trajectory) was significantly longer for aircraft in the Hard scenario than the Medium. While globally the average heart rate response was not found to be significant, several participants were measured to have more elevated heart rates during the Hard scenario as compared to the Medium. There was not a significant effect on heart rate response, the normalized taxiing time, and the number of stop and gos. Having experience working at Charles-de-Gaulle was also shown to be a significant covariate, on all of the same measures.

Participants were observed to have some difficulties managing the intensity of aircraft traffic when using paper flight strips. Additionally, information management on the radar screen posed some challenges in aircraft identification. These issues will be accounted for in the design of the MoTa platform.

Five operational events were introduced in these scenarios to allow for further exploration of the effect of automation on the types of problems air traffic controllers must face in addition to the primary task of directing aircraft. While all of the events were executed perfectly in the Medium scenario, the pilot error and closed taxiway events posed significant challenges in the Hard scenario. The other three (restricted taxiway, towed aircraft, change in configuration) were properly conducted in the Hard scenario. The performances of these five events over the two scenarios have been recorded and will be compared to those measured in experiments 2 and 3.

Experiments 2 and 3, testing the effect of the interface and the taxibots respectively, will occur in October 2015.

2 Introduction

2.1 Purpose of the document

This document describes the first experiment performed for Project Modern Taxiing (MoTa), the first of three to validate the proposed platform. The report starts with a brief review of the experimental design associated with the first experiment and focuses primarily on the results. It is assumed that the reader is familiar with Document E.02 24-MOTA-D3.1-Experimental Protocol [1], which discusses the entire validation campaign.

2.2 Structure of the document

Section 2 briefly explains the execution of the first experiment. Section 3 presents the analysis and the results. Section 4 presents the discussion and future work. Section 5 ends with a brief conclusion. Appendix A provides additional data more specific to individual performance.

2.3 Acronyms and Terminology

Term	**Definition**
A-SMGCS	Advanced Surface Movement Guidance and Control System
ATC	Air Traffic Control
ATCO	Air Traffic Controller
ATM	Air Traffic Management
CDG	Roissy Charles de Gaulle Airport
DIT	Deviation from ideal trajectory
E-ATMS	European Air Traffic Management System
GND	GROUND: ATC controlling position in charge of all the a/c from the block or gate to the runway and backwards
HRV	Heart Rate Variability
LF/HF	Low Frequency/High Frequency
MoTa	Modern Taxiing
NSG	Number of Stop and Gos
NTT	Normalized taxiing time
SA	Situation Awareness

Term	Definition
SART	Situation Awareness Rating Technique
TaxiBot	A aircraft tractor controlled by the pilot from the cockpit or fully automated that pulls aircraft on ground without using aircraft's engine power.
NASA TLX	NASA Task Load Index
PAC	Percentage of aircraft correctly treated
SESAR	Single European Sky ATM Research Programme
SJU	SESAR Joint Undertaking (Agency of the European Commission)
SJU Work Programme	The programme which addresses all activities of the SESAR Joint Undertaking Agency.
SESAR Programme	The programme which defines the Research and Development activities and Projects for the SJU.

3 Experiment

Experiment 1 featured one independent variable, Scenario complexity, which varied between Medium and Hard. Each experiment session lasted about 3 hours and contained four parts: Introduction and Practice; installation and calibration of neurophysiological equipment; run 1; and run 2. Participant feedback was collected after each run.

Prior to each session, the participant was sent some introductory material which included a discussion of the standard routes used at Roissy Charles de Gaulle (CDG). This information was based off of the CDG Operational Manual [2]. Additionally, a map of CDG south end was provided, and a suggestion of how to arrange the strip tableau. This information was also available to participants during the experiment session. The practice session featured a 20 minute scenario with significantly less aircraft, in order to introduce the participant to the simulator environment and the types of aircraft they would see. The session was guided and interactive, with a researcher explaining and answering participant questions.

After this practice session, the participants were equipped with an electrocardiogram, a single node electroencephalogram, and an eyetracker (Figure 1). The data for the latter two sensors will not be featured in this particular report and will be discussed in future reports. Baseline measures for the electrocardiogram and electroencephalogram for three minutes, and the eyetracker was calibrated using 12 points around the simulation environment (Figure 2).

Figure 1: (left) Electrocardiogram (not pictured), electroencephalogram, and eyetracker installed on a participant. (right) View from eyetracker forward-facing camera.

Figure 2: Eyetracker calibration points within the Simulation environment.

Each run was 35 minutes long, with the first either being Hard or Medium. This run order was balanced and also balanced with respect to experience working with CDG. After each run, workload and situation awareness measures were reported. Trust in automation was taken after the first run.

At the end of the experiment, the apparati were removed from the participant.

The experiment began in October 2014 and concluded in March 2015. Ten participants were completed in October / November, one in February, and one in March. This schedule reflects the availability of the participants.

3.1 Ground control interface

For this experiment, the simulated environment is Roissy-CDG ground control position in the current layout. As it can be seen in Figure 2, the experimental tower includes a simulated ground radar image close to the one used in Roissy-CDG (AVISO), a strip printer and a departure sequence manager (DISCUS).

AVISO is a more complete A-SMGCS system than the simulated version. After several workshops with ATCO and visits at Roissy-CDG south tower, the features that are essential for the ground controlled taxiing phase have been selected:

- color coded labels for arrival and departure a/c
- information displayed in the label : callsign, company, runway (for departure), parking (for arrival), a/c type
- parking zones and taxiways names, adjacent frequencies (for non CDG experienced ATCO)
- 1 second refresh rate of the radar positions

In the same manner, paper strip format currently in use at Roissy-CDG has been reproduced in the simulator and an actual strip printer is used. The simulation scenarios include precise and realistic timing and information to be printed on the strips.

These elements were sufficient, according to the interviewed ATCos, to emulate a ground control position that fulfills the experiment needs.

3.2 Scenario description

The Medium and Hard scenarios represented the range of controller taskload that would benefit the most from the MoTa platform. These scenarios were designed to be used specifically for the south end of CDG and are designed to be 35 minutes long. The Medium scenario features a maximum of 31 aircraft whereas the Hard scenario has a maximum of 51 aircraft.

Both scenarios included some simplifications from the full set of GND responsibilities at CDG. For example, the parking stand number was ignored. Participants only had to send the arrival to one entry taxiway per parking area. Similarly, departures always left from one exit taxiway per parking area. At CDG, some parking areas have several entry/exit taxiways and aircraft enter and exit depending on the parking stand. Ignoring the parking stand numbers reduced the GND's responsibilities for this project. Pushback at stands A30-33 was not required in the scenario (DGAC, 2008). Some parking areas were also merged and treated as one: A and C were combined (used the same entry/exit) and B and D.

Taxiway E was selected as the only area with a restriction. Both scenarios assume that there was good visibility. The traffic was composed of major international airlines and was approximately half francophone.

Each of these scenarios featured specific events that were designed to add task complexity (thereby increasing workload) to examine the potential advantages of the automation. Reference [1] provides a literature review and the methodology used to solicit the design of these operational events. There were four events in the Medium scenario, five in Hard: restricted area, pilot error, towed aircraft, taxiway closure, and solely for the Hard scenario, change in configuration. Table 1 and Figure 3 provide a summary of each event, including a representation of when the event occurs within the scenario.

Table 1: Summary of Operational Events For Each Scenario.

	Operational Events				
Scenario	**Config. Change**	**Closed Taxiway ***	**Pilot Error**	**Restricted Area**	**Towed Aircraft**
Medium	--	Dep. from F, stuck on RP15	Dep. from G, takes F instead of N	A380 Arr. from North, going to L	From North, to J
Hard	W to E, T+15 mins	Dep. from J, stuck on RP15	Arr. from North, takes F instead of N	A380 Arr. from North, going to L	From North, to M

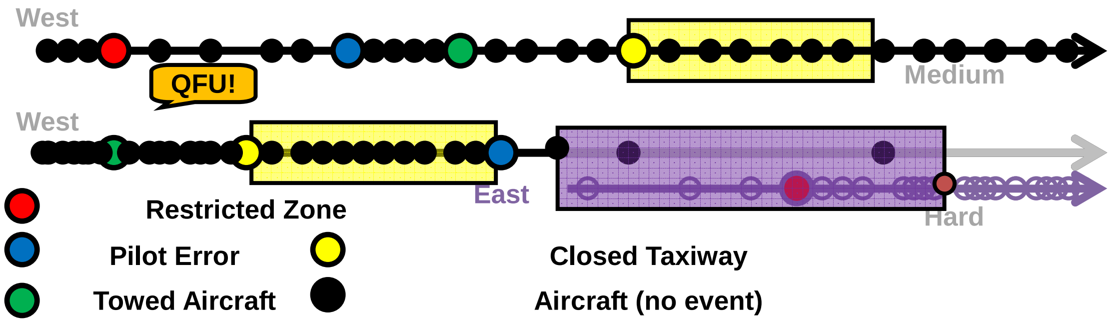

Figure 3: Timeline of Operational Events in each scenario.

The restricted area is defined as a taxiway that is closed off to a certain type of aircraft. In this case, we chose to simulate the restriction of part of taxiway E to the A380 due to its weight. Taxiway E extends between E5 and F and it is between taxiway B and F where this restriction exists. In both scenarios, an A380 arrival calls the tower from holding point Middle 2 and must park in L, with an entry taxiway of KL4. The shortest distance is through E. There is only one A380 in each scenario, but several aircraft in the Heavy category and several other Airbus aircraft. In both scenarios, the A380 is with Air France

The pilot error event is caused by a pilot taking the wrong taxiway, thus moving in counter-direction to the flow of traffic. The ATCO catches the error and redirects one of the aircraft or one of the two pilots (the mistaken or the oncoming traffic) calls the tower for further instructions. In the Medium scenario, this event occurs with a departure from parking G, taking taxiway F instead of taxiway N. In the Hard scenario, an arrival aircraft takes taxiway F instead of N. The pilot error event occurs near one of the FN* taxiways which connect F and N, which prevents the need for a tractor to pushback an aircraft.

The towed aircraft, always a B743, is pulled at half the speed of the other aircraft. This event originates in both scenarios at Middle 2. In the Medium scenario, the aircraft is being pulled to parking area J; in the Hard scenario, to parking M.

The taxiway closure, in both scenarios, occurs around taxiways E4, E5 and R. The affected Air France aircraft have a mechanical problem and are both departures (deemed to be more likely to have mechanical problems than an arrival that has finished its route without problem). In the Medium scenario, the aircraft is departing from parking F and blocks RP16/P1/E5. In the Hard scenario, the Air France aircraft is leaving from parking J and blocks RP15/P1/E4 (east configuration). The aircraft block the taxiway for five minutes.

The change in configuration only occurs in the Hard scenario and results in a change from the west configuration to the east. A warning is given to the ATCO within five minutes of the start of the scenario, with the warning noting that the new configuration will be active in ten minutes. Fifteen minutes after the start of the scenario, the new configuration is active. At the start of the new configuration (east), departures still in west must be rerouted. Participants were told of the last departure in the west configuration at the moment the east configuration was employed.

3.3 Simulator validation

3.3.1.1 Run order and pseudopilot

The covariates related to the execution of the scenario were tested for significance. The order in which the two scenarios were presented to the participant was counterbalanced. Additionally, a practice scenario was introduced at the beginning of the experiment to reduce any learning effects during the actual scenario runs. Run order was generally found to be not statistically significant using the Mann-Whitney U test, except for when participants self-reported working during the Hard scenario. In this situation, it has a mild effect (cf. Section 4.6).

All twelve scenarios featured the same two pseudopilots, with the third pseudopilot rotated based on scheduling availability. Seven scenarios were conducted with pseudopilot F, two with G, two with Z, and one with B. While all pilots were trained to competency, there was a concern that different performances were achieved. A Mann-Whitney U test was used to determine the effect. The pseudopilot covariate of two levels – pilot F, pilot non-F – was determined to not be significant on self-reported cognitive demand (see Section Y), and had a mildly significant effect on PAC ($U = 30$, $p < 0.0505$, $\widehat{PAC_F = 63}$, $\widehat{PAC_{non-F} = 78}$). However, this covariate also happened to be unintentionally confounded with CDG experience. In six of the seven runs performed by pseudopilot F, the participant was a non-CDG ATCO. In all five of the non-F pseudopilots, the participant was from CDG. It is more likely that the experience with CDG had a greater effect on the PAC than the effect due to the pseudopilot. Evaluating the performance change associated with each pseudopilot was not possible due to the small sample size. The pseudopilot group is not significant on the cognitive demand ($p < 0.7432$). Figure 4 illustrates the changes between groups for both the PAC and the cognitive demand.

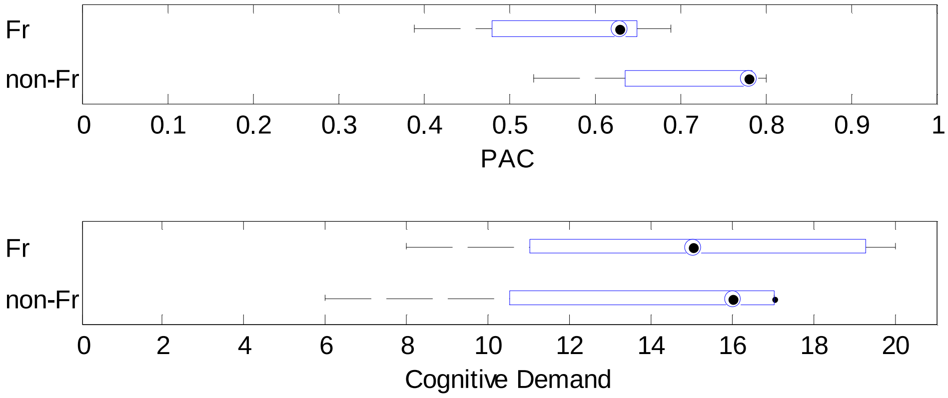

Figure 4: Effect of pseudopilot group on the Scenario execution.

3.3.1.2 Quantitative and qualitative participant feedback

A questionnaire was distributed to the participants *post-hoc* for their assessment of the simulator fidelity. Of the 12 participants, six responded, providing a quantitative score of this simulator and the reference simulator associated with their home airport. The questionnaire featured eight statements with a 5-point Likert scale, one for each construct such as verification, conceptual validity, and internal validity (cf. [ATCpaper]). Two questions were reserved for open-ended feedback on the likes and dislikes of the simulator fidelity. Reference [4] discusses the development of this questionnaire in greater detail.

On average, the six participants rated our simulator 3.46 out of a maximum of 5 points ($\sigma = 0.74$, min score = 2.75, max score = 4.88). Participants generally appreciated the physical work environment and the realism of aircraft movements. The external view was well appreciated by some, while others noted that the entire 360° view was not fully modelled. Participants also wished for a larger or additional strip board. Participants stated that the traffic load was realistic and the aircraft reacted quickly to the participant's commands. However, for some participants, the aircraft did not react quickly enough, thus reducing the participant's ability to predict upcoming traffic movements.

Participants did not like the fact that the other controllers were not simulated and the teamwork between each of the roles was not explored, even though that was not the point of our research.

In general, participants felt that our simulator was realistic enough to evaluate the project goals ($\mu_{valid} = 4.33, \sigma_{valid} = 0.52$) and accurately reflected user's experience and skill ($\mu_{crit} = 4.00, \sigma_{crit} = 0.63$) but that it lacked conceptual validity ($\mu_{concept} = 3.00, \sigma_{concept} = 1.41$). Several participants noted that if approached with the same scenario in the real-world, they would not react the same way ($\mu_{conv} = 3.00, \sigma_{conv} = 1.67$). Overall, the participants were slightly in agreement on the applicability of the results to both CDG and other airports, with questions 7a and 7b[1] scoring an average of 3.17 ($\sigma_{7a} = 1.17$) and 3.33 ($\sigma_{7b} = 0.82$) respectively. This feedback provides reassurance regarding our choice of using CDG as the paradigm for testing new automated taxiing technologies. Figure 5 illustrates the range of scores across all of the dimensions.

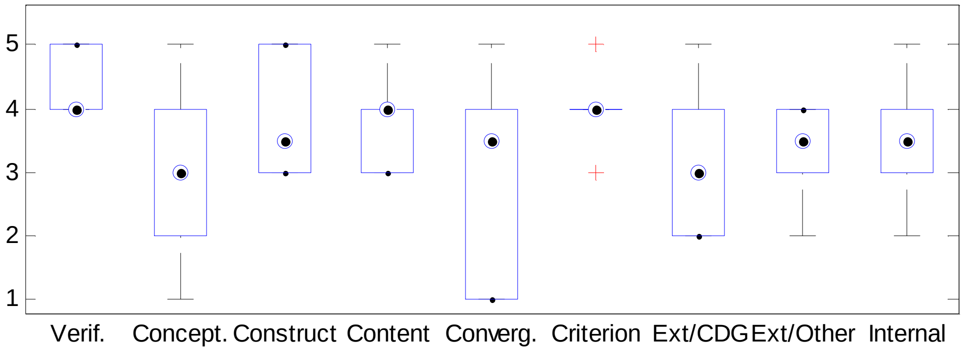

Figure 5: Scores across simulation validation dimensions.

As a comparison point, the participants were also asked to rate the simulator of their home airport. On average, the three participants who rated the CDG simulator gave it a score of 4.75 of 5 points, $\sigma = 0.125,$ min score = 4.63, max score = 4.88). They noted that simulator was well-adapted to the actual work environment (albeit with changes, such as the use of a mouse for the radar screen, or non-simulated buildings) and allowed for multiple controllers and roles, including interactions with other members of the team such as the Tower Supervisor, firefighters, the ATCOs handling final approach, etc. This subset of participants rated our simulator, on average, 4.04 points. Two participants provided ratings for the simulator at ENAC, which scored an average of 3.44 points (min score = 3, max score = 3.875). These same participants rated our simulator 3.31 points. The response for Orly was ignored due to the singular data point.

1

[1] 7a: The Simulator is realistic enough that results and trends from the User performance of the ground controller conducted in the Simulator could be extrapolated to real world performance at Roissy Charles-de-Gaulle.7b: The Simulator is realistic enough **and the representative situations are sufficiently comprehensive** such that the results and trends from the User performance of the ground controller conducted in the Simulator could give an estimation of the real-world performance **of new technology concepts at any airport**.

4 Results

There were 13 participants (2 women), with 12 persons’ data points used. One person was eliminated due to simulator difficulties; there were technical problems during the execution of the Hard scenario for him. The average age of all participants is 41.8 years old (std = 10 years), with a range of 35-68 years. On average, the participants had 13.4 years of ATCO experience (std = 8.43 years, range of 6-35 years), with all participants having performed the role of the ground controller. There were 5 participants from Roissy Charles-de-Gaulle airport, 3 from Paris Orly Airport, 1 from Melun Villaroche Aerodrome, 1 from Goteborg Landvetter Airport, 1 from Strasbourg International Airport, and 1 from Toulouse–Blagnac Airport.

In this initial experiment, the only independent variable is the scenario complexity (Medium, Hard). There are several dependent variables: the percentage of aircraft correctly treated (PAC), deviation from ideal trajectory (DIT), normalized taxiing time (NTT), number of stop-and-gos (NSG), heart rate variation (HRV), self-reported workload in the form of NASA Task Load Index (TLX) [5], self-reported situation awareness (SA) in the form of the Situation Awareness Rating Technique (SART) [6], SHAPE Automation Tool Trust Index (SATI) [7]. There were also three covariates: run order (Average then Hard, Hard then Average), experience working at CDG, and pseudopilot (F, non-F). Each subsection defines the individual variable and describes the possible interpretations that could be made, including caveats. It is then followed by the raw statistical results. The next section provides an overall discussion of the results.

Variable	Units	Definition	Trend	Caveats
Percentage of Aircraft Correctly treated (PAC)	N/A	Ratio of aircraft correctly treated to the maximum number of aircraft possible for the scenario	Larger PAC indicates stronger performance	Does not directly provide number of aircraft treated in 35 minutes
Deviation from Ideal Trajectory (DIT)	Time, mins	Taxiing time exceeding the minimum taxiing time required to traverse the shortest trajectory while following standard airport routes	Smaller and negative DITs indicate less bottlenecks, more efficient routes (e.g. shortcuts).	Not possible to separate the source of deviation (bottlenecks, non-standard routes)
Normalized Taxiing Time (NTT)	N/A	Ratio of actual taxiing time to taxiing time of the ideal trajectory (no traffic, shortest path while following standard airport routes)	NTT of 1 is equivalent to a DIT of 0; an NTT less than or equal to 1 indicates better performance	Not possible to separate the source of deviation (bottlenecks, non-standard routes)
Number of Stop and Gos (NSG)	count	Number of times an aircraft stopped and continued its route	Less indicates more efficient aircraft taxiing	Related, but not equivalent, to number of bottlenecks
Heart Rate Variation	Beats per	Change in the Average number of beats per	Larger HRV indicates greater	Not possible to separate the source of deviation

(HRV)	minute, bpm	minute from a baseline sample	stress, implying larger workload	(taskload)
Task Load Index (TLX)	N/A	Self-reported workload using NASA's methodology	Larger values indicate more stress	Self-reported post hoc, strong correlation to performance
Situation Awareness Rating Technique (SART)	N/A	Self-reported situation awareness using one methodology	Larger values indicate stronger situation awareness	Self-reported post-hoc, strong correlation to performance, non-zero center
SATI	N/A	Self-report trust in automation using Eurocontrol's methodology	Larger values indicate more trust in automation	Self-reported post-hoc, on technology that was not used during simulations
Run Order	N/A	Two possibilities: Hard then Medium; or Medium then Hard	n/a	Learning effects are prevalent, especially in non-CDG ATCOs
CDG Experience	Yes, No	Experience working at CDG		Does not differentiate between airport sizes
Pseudopilot	Fr, Non-Fr	Third pseudopilot for the Hard scenario, either the standard (Fr) or a substitute (non-Fr)		Assumes that the individual pseudopilot variation across participants is the same

Due to the small sample size, a Wilcoxon signed-rank test was used to determine the effect of scenario complexity, as the samples between the two levels are paired. A Mann-Whitney U test was used for comparisons between ATCOs with and without CDG experience.

4.1 Percentage of aircraft correctly treated

Each scenario is designed with a maximum number of vehicles, as noted in Section 3.2. An aircraft is considered as correctly treated if it has been successfully transferred to the consequent sector (local or apron) by the participant. This transfer point is the last point of contact with an aircraft, after the initial call and any follow-up commands. The PAC is calculated by taking the number of correctly treated aircraft within the 35 minute scenario and dividing by the maximum possible. Aircraft that are in mid-route at the 35 minute mark are counted as correctly treated, as we assume that a transfer would occur at the end of this route. Pseudopilots are trained to prompt the participant if the transfer has not been initiated prior to arrival at the sector boundary, and are told to respect the timing of the scenario to their best of their ability until after the 35 minute mark.

The variable is the strongest indicator of overall performance and correlates directly to a measure that is used in real-life applications. Large PACs indicate stronger performances whereas smaller PACs indicate weaker performances.

Scenario difficulty was determined to be significant (Z = 3.060, p < 0.0004), with a $PAC_{Med} = 95.3$ and $PAC_{Hard} = 62.9$. Globally, having CDG experience has a mildly significant effect on PAC (W = 41, p < 0.09235, $\widetilde{x}_{non-CDG} = 0.745$, $\widetilde{x}_{CDG} = 0.885$). However, there appears to be an interaction between the two variables. For those without CDG experience, the scenarios were significantly different (Z = 2.3707, p < 0.01562, means of 0.923, 0.547 for Medium and Hard, respectively) and mildly significant from ATCOs from CDG (Z = 2.0226, p < 0.0625, means of 0.994, 0.744 for Medium and Hard, respectively). Figure 6 illustrates this effect.

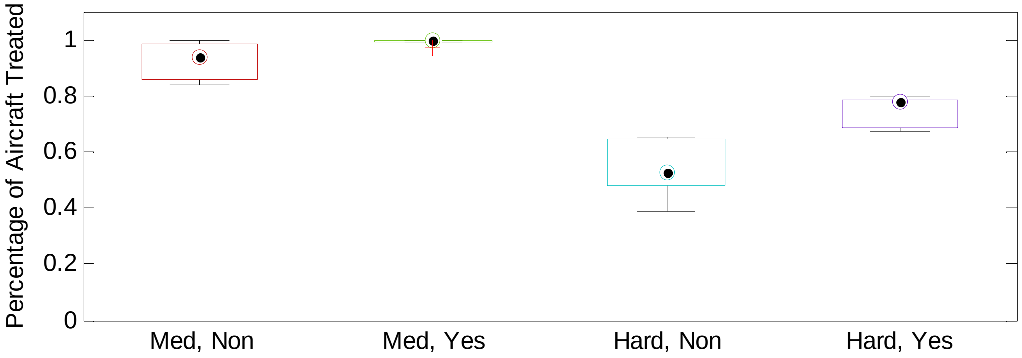

Figure 6: Effect of Scenario and CDG experience on Percentage of aircraft correctly treated

4.2 Deviation from the ideal trajectory

The total taxiing time is, in theory, the sum of several factors: the time to traverse between the start and end points with respect to the vehicle’s velocity and the deceleration or stoppage time due to yielding or stopping behind an aircraft. The total taxiing time is highly dependent on the sample fleet’s demographics; i.e. the type of aircraft and the intended distance. Therefore, results can be misleading. For example, the average taxiing time of a homogeneous fleet traversing a small distance would be substantially less than of a mixed fleet traversing a long distance. As such, we introduced the deviation from the ideal trajectory to explain part of the story.

The ideal trajectory was defined as the shortest physical distance between a starting and final point based on standard airport rules (i.e. no taxiing against the regular flow of traffic) with no traffic in front of the particular aircraft. This trajectory was calculated for every aircraft planned in the simulation, keeping into account the following assumptions:

- The taxiway W9 is used to access the runway in the west configuration
- The taxiway S2 is used to access the runway in the east configuration
- The ideal trajectory for the planned configuration is used, meaning that rerouted aircraft or west-bound aircraft that are taxiing in the east configuration are counted against the west configuration.
- It is difficult to retrieve the minimal trajectory time with respect to the trajectory actually travelled, as there was a great number of variations between participants.

This value only accounts for the time within the ground sector, and does not factor in the initial call. Based on these assumptions, it is possible that a minimal trajectory time with respect to the actual trajectory would provide differing results. A trajectory time meeting exactly the ideal would result in a value of 0. It is possible to have negative or a positive value, with the units being minutes.

Scenario complexity was determined to be significant (Z = -2.981, p < 0.0009). CDG experience is also significant with respect to scenario complexity, but only for those without CDG experience (Z = -2.3664, p < 0.0156). Figure 7 illustrates this effect on DIT.

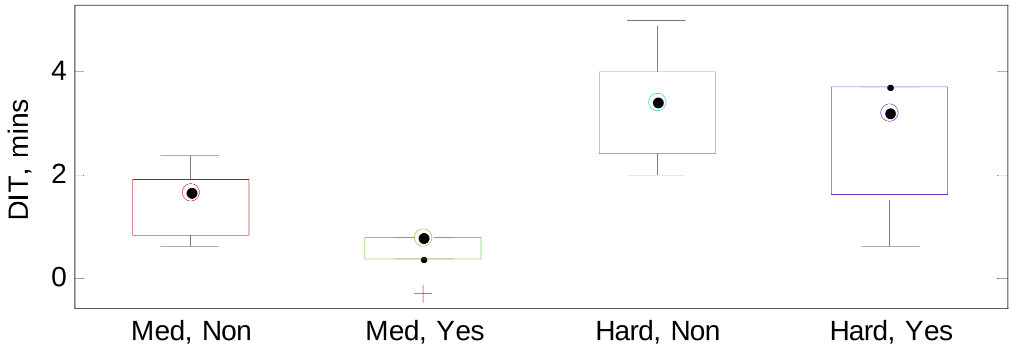

Figure 7: Effect of Scenario complexity and CDG experience on Deviation from Ideal Trajectory.

These values indicate that aircraft managed by participants without CDG experience are taxiing longer than expected. However, it is not clear as to the source(s) of the delay. There are several contributions, within the context of this simulation: the route clearance, the traffic congestion, and the method for giving the command. Participants may not have been aware or unwilling to take shortcuts (e.g. going across the standard flow of traffic). An aircraft may have yielded to traffic or followed slower aircraft. Participants may have opted to provide pilots part of the route clearance (‘hold before R’) and then provided additional guidance. However, if new directions are not given prior to the intended intermediary point, the pilot must stop and call the tower for further commands, thus leading to more delays. The assumption is also made that errors due to pseudopilot control do not change between participants.

4.3 Normalized Taxiing Time

We also sought to determine a normalized version of the taxiing time, based on the ideal trajectory. This variable allows for comparison between aircraft and scenarios. The normalized taxiing time is the ratio of the actual taxiing time to the taxiing time of the ideal trajectory, as defined in the previous section. An NTT of 1 is the equivalent to a DIT of 0. Less than 1 would indicate that the taxiing time was less than the ideal trajectory time (DIT < 0). Similarly to DIT, the NTT has the same sources of deviation.

Scenario complexity was not determined to be significant (p < 0.9546), however, CDG experience had a mildly significant effect on this value (W = 100, p < 0.08397, Figure 8). Results indicate that participants with CDG experience, on average, had a lower normalized taxiing time than those without this experience.

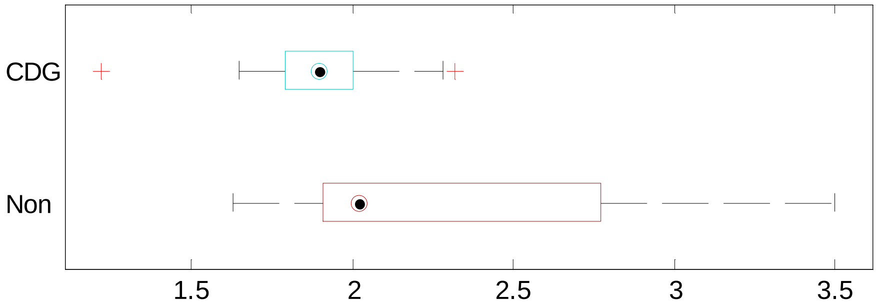

Figure 8: Effect of CDG experience on average normalized taxiing time.

4.4 Number of Stop and Gos

This variable is defined as the number of times an aircraft stopped and continued its route while in the ground sector. A lower number means that the aircraft stopped less frequently – also lowering its DIT and NTT. Since it was not possible to accurately count the number of bottlenecks in the scenario, the NSG was used as an approximation. The value exists only within the context of this simulation. Since the pseudopilots cannot regulate aircraft speed, they must stop with respect to the distance between aircraft. Therefore, if an aircraft is stuck behind a slower aircraft, it would frequently stop and go while to avoid traversing through the slower aircraft. Since this type of stop of go is similar to engaging the break, the NSG provides an estimation for a bottleneck.

However, it is difficult to draw direct parallels. Several aircraft behind a slow moving aircraft would result in a higher NSG for those aircraft, but it stems from a single bottleneck. One may argue that the NSG provides the significance of the bottleneck (more affected aircraft), but it is hard to isolate the specific moment or the cause. Additionally, stop and gos may be the result of an ATCO's route clearances. Stopping at an intermediary holding point would increase the NSG.

In this case, the average NSG per aircraft was calculated for each participant and compared. Scenario complexity was not determined to have a significant effect ($p < 0.3535$, Figure 9).

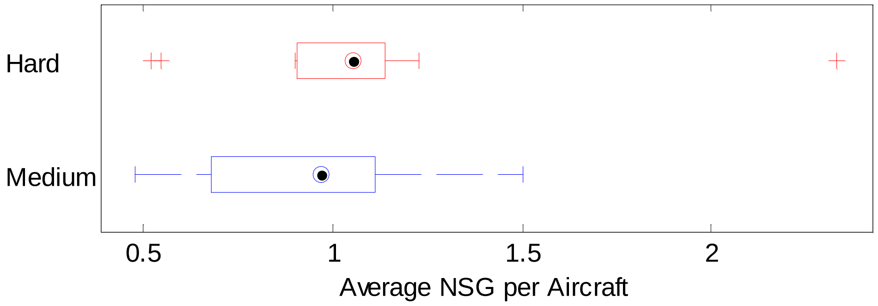

Figure 9: Effect of Scenario on Average NSG per aircraft.

4.5 Heart rate response

Heart rate response is calculated by subtracting the average heart rate measured during a three minute baseline from the average heart rate measured during the scenario. When plotted against time, this response gives the participant's physical response to the scenario. The raw heart rate data was collected at 512 Hz and was processed and filtered using Kubios HRV [8]. Figure 10 illustrates an example heart rate response profile for one participant. The data shown is the moving average of the heart rate in beats per minute (bpm), with a window of 0.0625 s. In this particular case, the participant had greater changes in heart rate during the Hard scenario than the Medium scenario.

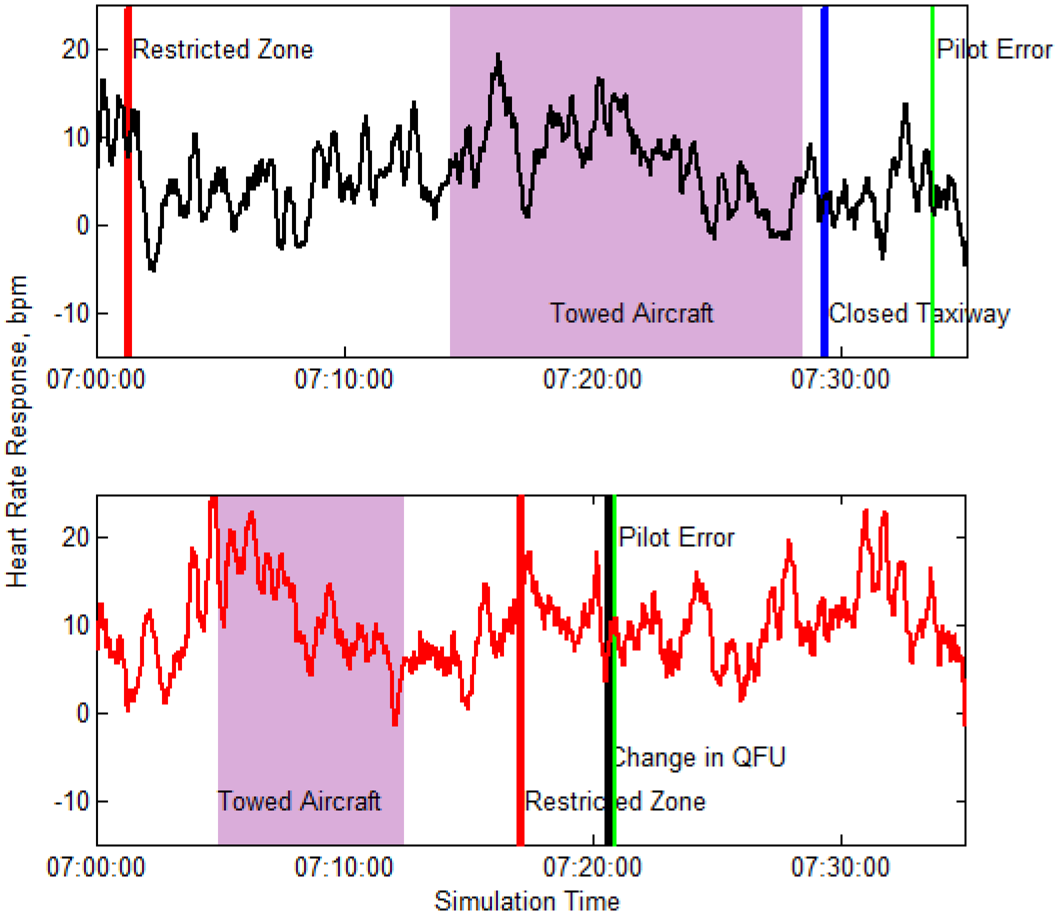

Figure 10: Exemplary heart rate response profile for one participant.

Neither scenario difficulty nor experience at CDG has a significant effect on this measure ($p < 0.2383$ and $p < 0.9697$, respectfully; Figure 11). The LF/HF ratio (a proxy of workload) was also measured for the first five minutes of each scenario (Figure 12). Neither scenario difficulty, CDG experience, nor run order had a significant effect on this measure (lowest p-value at 0.14).

HR response was divided into two groups based on TLX score, with the median TLX score (4.585) as the divider. These two groups were compared to determine if there was a significant difference between the HR response of those who gave a low TLX (TLX > 4.585) score to those who gave a high TLX score (TLX > 4.585). The means of these two groups differed ($\mu_{low\,TLX} = 4.23$, $\mu_{high\,TLX} = 4.50$), however, there is no significant difference ($Z = 0.1569$, $p < 0.9097$). This result was further confirmed by evaluating the Pearson's correlation coefficient, which is 0.107 ($p < 0.6178$) between the TLX and HR response.

While several participants were measured to have a greater variation in heart rate when faced with the Hard scenario as compared to the Medium scenario, this result was not entirely conclusive. It is likely that the duration of the scenario (35 minutes) has a diminishing effect on the magnitude of the average heart rate response.

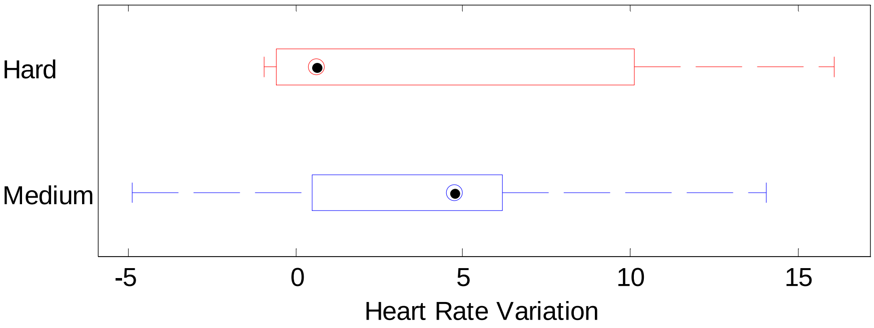

Figure 11: Effect of scenario on overall HRV.

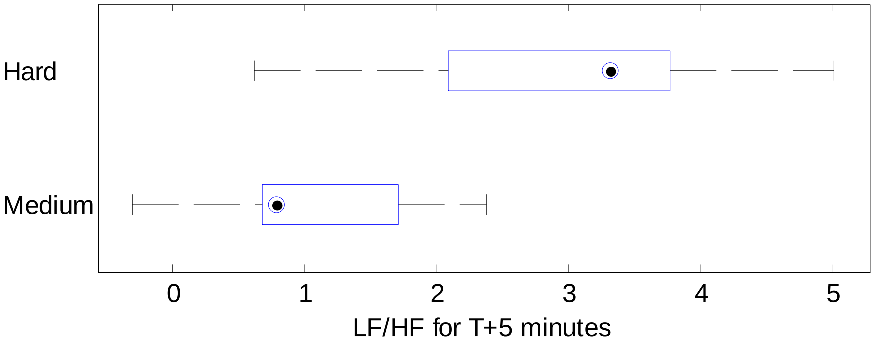

Figure 12: Effect of scenario on LF/HF for the first minutes.

4.6 Workload

Workload was self-reported by the participants at the end of each scenario with the NASA TLX questionnaire. The simplified version of the questionnaire was used for this experience, with no inter-dimensional ranking. Participants gave a score between 1 (low) and 7 (high) for each of the seven dimensions. The final TLX score was determined by averaging all values. The final scores range from 1 to 7.

Scenario complexity was a significant effect on the TLX score (Z = -2.9413, p < 0.001, $\widetilde{TLX}_{Hard} = 5.585$, $\widetilde{TLX}_{Med} = 3.83$). The Hard scenario was determined by participants to induce more workload than the Medium scenario. Experience at CDG was also determined to be significant (Z = -3.0653, p < 0.0048, $\widetilde{TLX}_{CDG} = 3.58$, $\widetilde{TLX}_{non-CDG} = 4.92$). Naturally, participants with experience at CDG found the scenarios to be less difficult than those who are unfamiliar with this airport. The effect of both of these variables is presented in Figure 13. Run order was mildly significant on the TLX scores during the Hard scenario (U = 30, p < 0.0501), with a higher score given if the Hard scenario was presented first (Figure 14). However, this effect is not significant in the Medium scenario. Figure 15 illustrates the distribution of scores for all of the dimensions.

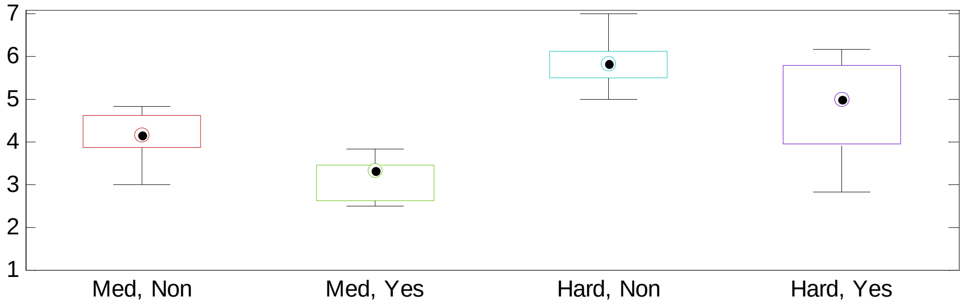

Figure 13: Effect of Scenario and CDG experience on overall TLX score.

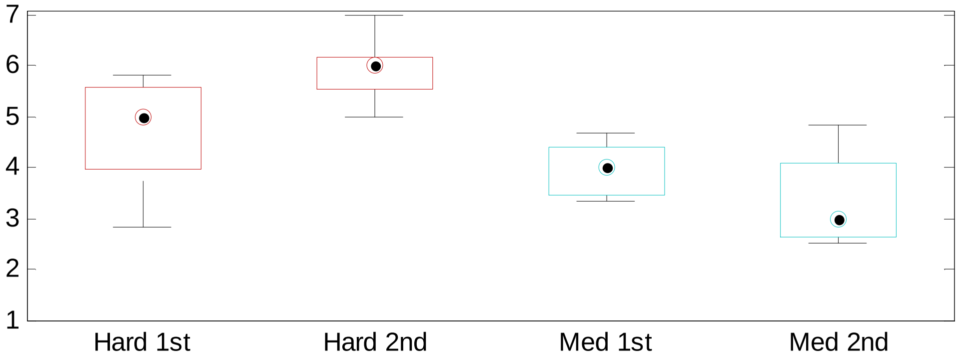

Figure 14: Effect of run order on overall TLX score.

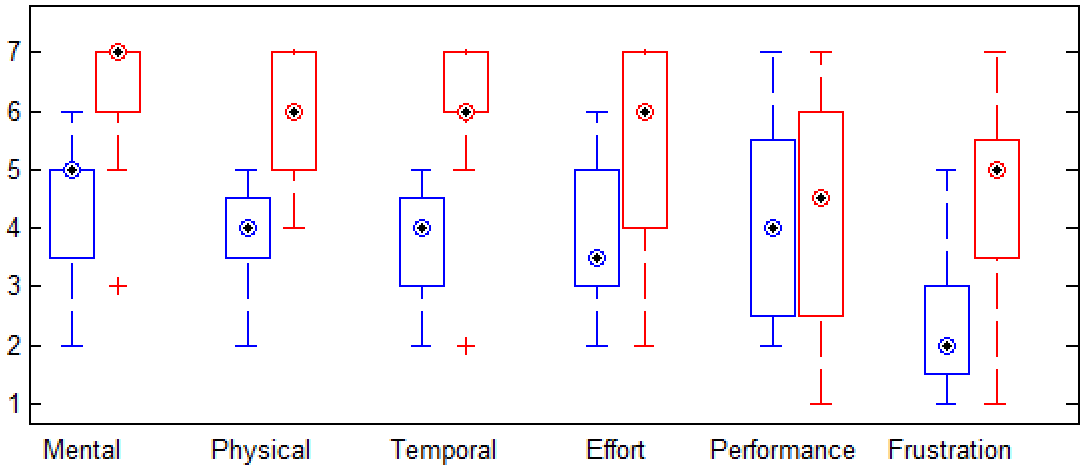

Figure 15: Range of TLX scores based on the six dimensions.

4.7 Trust in Automation

Trust in automation was self-reported by the participants at the end of the experiment with the SATI questionnaire. The participants filled out the survey based on the technology in their home airport, and not the experiment development. This assumption was made in order to eliminate any potential bias based on the simulation itself, which could possibly lead to a misleading acceptance of the proposed technologies in experiments 2 and 3. Figure 16 summaries the distribution of scores for each dimension. The black dot inside a circle is the median, the box marks the 25th and 75th percentiles, and the whiskers are to the largest and smallest non-outlier data points.

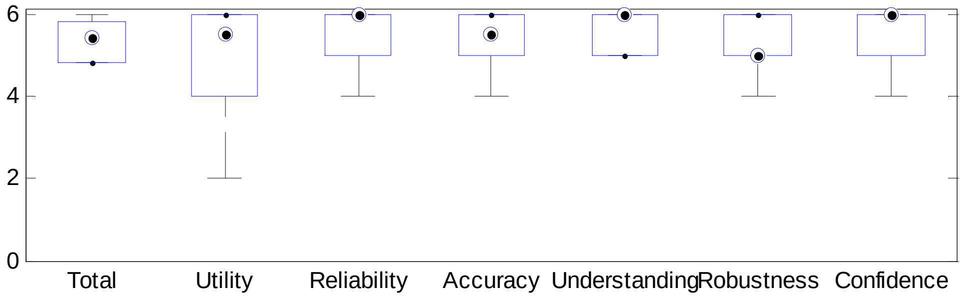

Figure 16: Distribution of SATI scores across all dimensions.

In general, the scores indicated a high acceptance in the current technology present at French airports, with an average score of 5.35 out of 6 (σ = 0.49, 0 being little trust, 6 being highest trust). Of the six dimensions, participants disagreed least with the statement that the use of the paper strips and the ground radar screen was useful, however, it was still globally ranked high (average score of 5, σ = 1.33). However, they did feel that this system was reliable (μ = 5.5, σ = 0.71) and had confidence using this system (μ = 5.5, σ = 0.71).

4.8 Situation awareness

Situation awareness was self-reported by the participants at the end of each scenario using SART. This questionnaire is administered similarly to the NASA TLX, in that participants rank between 1 (low) and high (7) over ten dimensions. The ten dimensions are then grouped into three categories: cognitive understanding, cognitive demand, and cognitive supply. Understanding is akin to a property of the individual, a characteristic or parameter that is set prior to the specific scenario, the comprehension and experience with the scenario and the task. Demand is the product of the scenario and can be likened to a requirement for execution of the scenario. Supply is similar to a resource of the participant. Thus, the final SART score is calculated as the subtraction of the difference of Supply from Demand from Understanding.

Scenario complexity was significant for cognitive demand (Z = -2.4162, p < 0.015), with the Hard scenario noted as requiring more situation awareness than the Medium scenario (Figure 17).

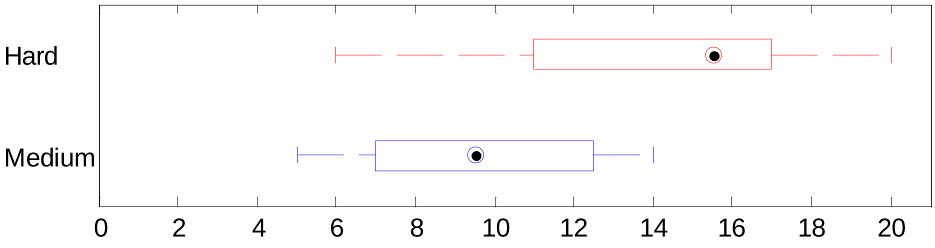

Figure 17: Effect of Scenario on Cognitive Demand

The maximum and minimum situation awareness score possible do not fall on a natural scale, that is, the minimum score possible is not 0, but instead -14. This score assumes all 1s for the dimensions under the Understanding and the Supply categories, and all 7s for Demand. The highest possible score is 49, assuming an inverse in scores as the previous example. Scenario complexity was determined to be significant on the situation awareness ratings (Z=-2.2706, p < 0.021), indicating that participants reported having less situation awareness during the Hard scenario than in the Medium (Figure 18).

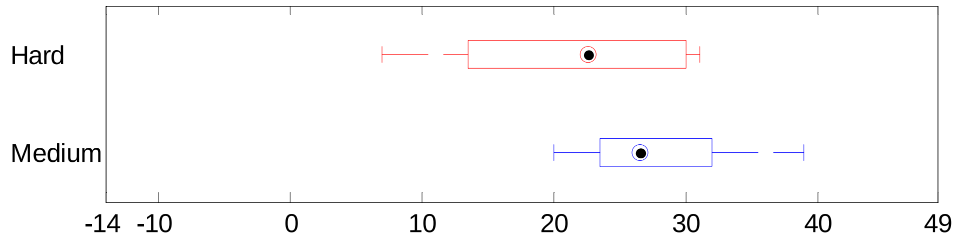

Figure 18: Effect of Scenario on overall SART score

4.9 Events

4.9.1.1 Execution

This project is unique as it is the first (to the author's knowledge) to use operational events such as these. Other projects have previously used the change in configuration, but none have attempted to isolate specific activities beyond the standard rerouting of arrivals and departures. While these events provide multidimensionality to the analysis, they are also difficult to execute and in a manner that is equitable between participants. Additionally, participant behaviour was a significant impact on the execution of the events – selection of a non-normal path, delays, and so forth. If in the event of a missed pre-planned event, the pseudopilots were asked to improvise the event later in the scenario. In some cases, particularly during the Hard scenario, adding the event prior to the end of the scenario was too difficult to achieve, thus, not all participants saw all events. Table 2 summarizes which events were missing during the hard scenario. Green indicates that event occurred at some point during the scenario, red means event did not occur.

Pilot error and closed taxiway were often the two events that were most missed. There are several explanations. To avoid predictability between the two scenarios, the pilot event was associated with an arrival in the Hard scenario instead of a departure in Medium. However, it is more difficult to arrange a face-to-face conflict using an arrival than a departure, as a departure has more flexibility in terms of entry into the ground sector than an arrival (a parking exit taxiway instead of waiting at S1-3 or S7/9 right off of the runways). A closed taxiway was also difficult to execute as it required greater attention from the pseudopilot and precision to ensure that the aircraft stopped where necessary. By far the greatest determinant in the execution of each scenario is the pilot performance – delayed execution and treatment of aircraft would often delay the entire schedule by several minutes.

Table 2: Frequency of operational events in the Hard scenario.

Part	sc	Restricted Zone	Pilot Error	Towed Aircraft	Closed Taxiway	QFU change
CARFL	**h**	🟩	🟥	🟩	🟥	🟩
DOVLO	**h**	🟩	🟩	🟩	🟥	🟩
GUSTO	**h**	🟩	🟩	🟩	🟥	🟩
MAIJU	**h**	🟩	🟥	🟩	🟩	🟩
MESCH	**h**	🟥	🟥	🟩	🟥	🟩
PANBR	**h**	🟩	🟥	🟩	🟥	🟩
PASLI	**h**	🟩	🟥	🟩	🟩	🟩
PIEAN	**h**	🟩	🟩	🟩	🟩	🟩
RASFR	**h**	🟩	🟥	🟩	🟥	🟩
SAIJU	**h**	🟩	🟩	🟩	🟥	🟩

SANPH	**h**	🟩	🟥	🟩	🟥	🟩
ZIMOL	**h**	🟩	🟩	🟩	🟩	🟩

These execution difficulties and the missing data make it difficult to compare between scenarios the individual performance during each event type. Indeed, the original purpose of collecting this type of data was to analyze the change in performance between experiments. As such, the results listed below do not draw any conclusions between scenarios and strictly quantitative and qualitative summations of the events that happened. Any variation between individuals for each event is treated as noise that would naturally occur.

Nevertheless, the lack of data could be treated as a variable itself. One could theorize that improved performance (i.e. as seen in the Medium compared to Hard) would prompt the arrival of such events. In theory, Cochran's test could be used to compare between the three experiments, as seen in Table 3 (pilot error used in the example). In this example, if another 24 experiments were used (12 for each experiment) and pilot error was executed in 23 out of 24 cases in Exp 2 and in all cases in Exp 3, a Cochran's Q of 10.3 ($p < 0.0057$). If this case were to happen, it would be evident that Experiments 2 and 3 are different from 1, thus implying that potentially the automation proposed in each of these cases made a significant contribution.

Table 3: Example of event frequency count analysis.

Occurrence	**Exp 1**	**Exp 2**	**Exp 3**
Yes	18	23	24
No	6	1	0

4.9.1.2 Restricted area

The restricted area event (Table 4) occurred in 23 out of 24 cases, with 11 out of 12 participants seeing it in the Hard scenario. Of these 23 instances, two participants tried to send the A380 through E, and three started to send it through E, but later caught their own errors and corrected it prior to the pilot calling back. One of the participants that corrected his own error had experience with CDG; the other four that did not initially send the aircraft through E were not from CDG. This rule is still active at CDG today.

Table 4: Number of occurrences for the Restricted Area Event.

Outcomes of the Restricted Area Event	**# of instances**
Did not occur in run	1
Did occur, sent through E and pilot called back	2
Did occur, sent through E but self-corrected	3
Did occur, participant did not send through E	20

The taxiing time of each of the A380s (AF649Z in Hard, AFR007 in Medium) was normalized to their ideal trajectories. Twenty-one cases were used as for one participant this same aircraft was used for the closed taxiway event and another had a software issue. On average, these aircraft were not significantly delayed, with a normalized taxiing time of 1.07. The Hard average (10 data points) was 1.13 (an additional 1.17 minutes) and for Medium (11 data points), 1.02 (an additional 0.18 minutes, or 11 seconds). It should be noted that these delays are not necessarily due to only the stoppage, but are an accumulation of all sources (additional traffic, route variations, etc). Figure 19 summaries the normalized taxiing time for each participant across both scenarios.

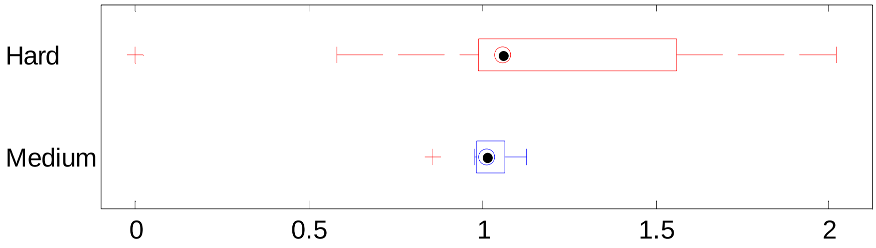

Figure 19: Normalized taxiing time for the A380 in the Restricted Area event.

4.9.1.3 Pilot Error

The pilot error event occurred in 17 out of 24 events, with half of the participants seeing it in the Hard scenario. In the Medium scenario, this conflict happened between FIN874P and JAT240 except for two situations. One participant sent FIN8749 through E in lieu of R; the event was later improvised later during the scenario between EZY808H and EZY343T on taxiway N (both are effectively replicating the previous routes). In another situation, the conflict occurred between AF689VC and FIN874P. In the Hard scenario, this pilot error occurs when AF795JZ, an arrival, takes taxiway F instead of N and finds himself face to face with AFR275. Figure 20 summarizes the routes prior to the moment of collision.

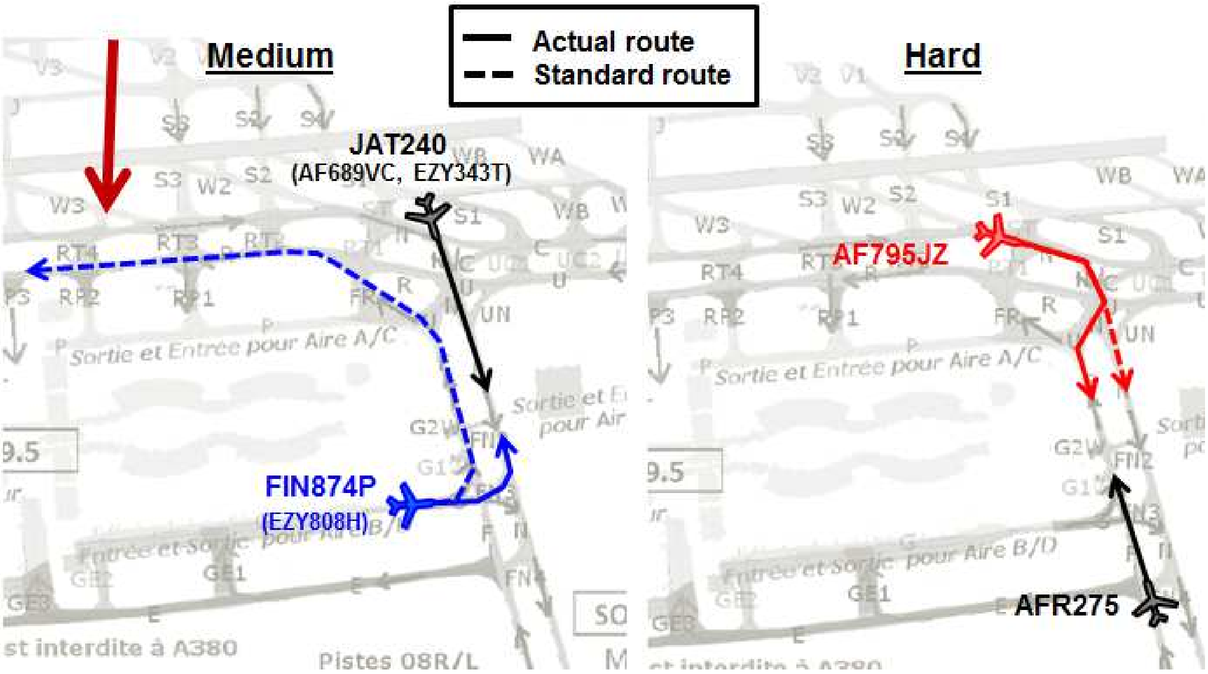

Figure 20: Visualization of the Pilot Error event in both the Medium and Hard scenarios.

At the moment of collision identification, the participant needed to determine which of the two aircraft needed to be rerouted and how. The event was designed for at least one solution, with the most obvious being a short deviation via FN2. In 14 out of 17 cases, the participant chose to reroute the offending aircraft. It is quite possible that in the future experiments, the automation may end up rerouting the aircraft on an entirely different route that accounts for the current traffic. The normalized trajectories of the two aircraft involved in the pilot error were averaged for each participant, and then averaged across both scenarios for an idea of the trajectory deviation. In general, the normalized taxiing time with respect to this event was 1.59. It is important to note that this value includes possible delays later in the trajectory, after the pilot error event.

The reaction time was measured from the start of the taxiing time from the offending aircraft until the moment that either the pilot or the participant reacted to the event. This reaction was defined as a call from the pilot or a verbal comment as captured by the audio recorded (examples include: “what is the Finnair doing?”, “Finnair, you’ve made a wrong turn”, “JAT240, hold position”). The resolution time was defined as the moment of the reaction to the end of the command given by the participant. This time accounts for any deliberation time and time to send the command. In 7 cases, the ATCO noticed that there was a deviation in the trajectory and rectified the situation. In 10 cases, one of the pilots (either the one taking the wrong trajectory or the one affected by the mistaken pilot) called the tower and asked for further directions. In all cases, there was at least thirty seconds possible for the participant to identify the mistaken turn.

On average, the reaction time was relatively rapid, with participants reacting within 41 seconds (32 s in Medium, 60 s in Hard). The resolution time was much faster, with an average of 37 seconds (44s and 23s in Medium and Hard, respectively). Figure 21 summarizes the reaction times for each participant. It is worth noting that only five data points for the Hard scenarios.

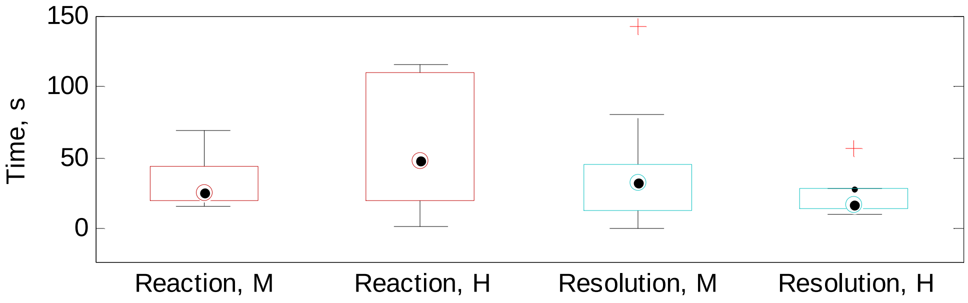

Figure 21: Reaction and Resolution times for the Pilot Error event.

4.9.1.4 Towed aircraft

A towed aircraft event occurred in all runs for all participants (24 cases). The two tractors were HTR and SDH, respectively, for the Medium and Hard scenario. In the Medium scenario, two routes were possible: through E and through R. In the Hard scenario, the most direct path was through F. Figure 22 illustrates the path taken by each of these tractors. The participants were not given any special directions on how to handle the tractor, aside from the fact that the tractor moved at a velocity of half of the other aircraft. Therefore, each participant had the choice to delay or reroute the aircraft as desired, with no direct consequence.

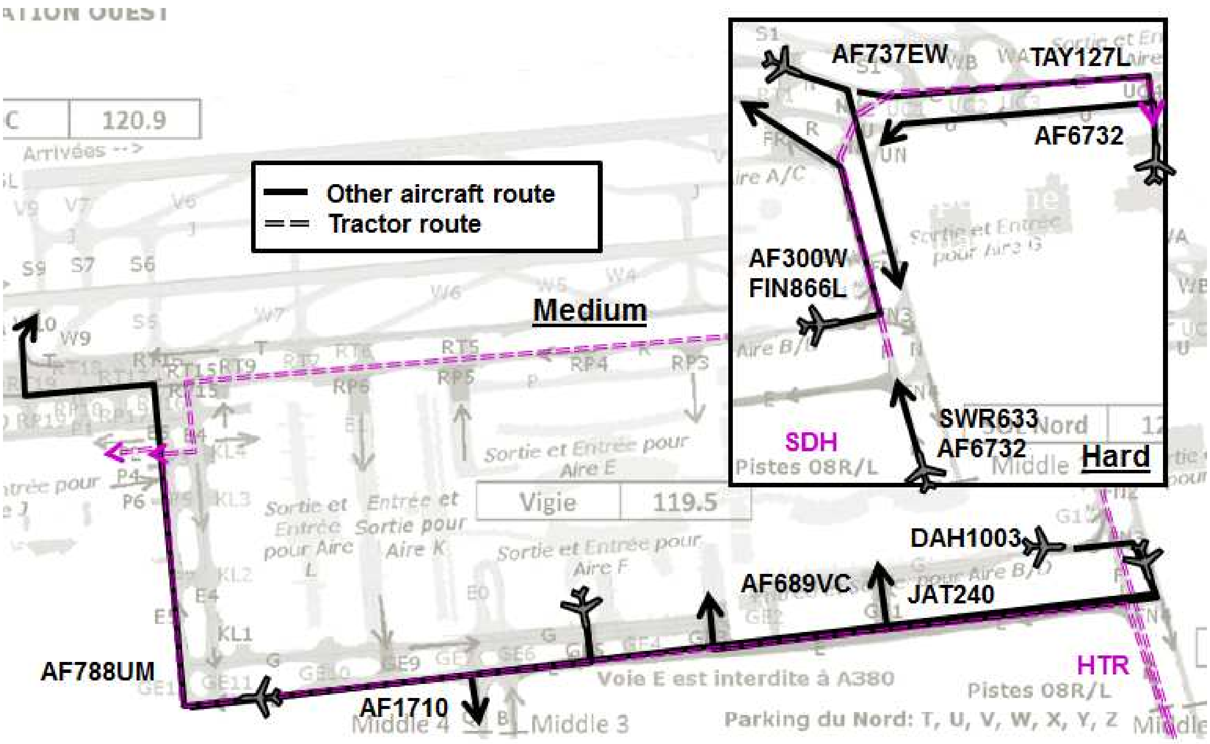

Figure 22: Visual representation of the Towed Aircraft event.

The taxiing time of the tractor and the other aircraft affected by it was recorded, normalized, and averaged for each participant. The other aircraft affected were noted as those that yielded or were yielded to by the tractor itself or those that had entered the ground sector at the same time as the tractor movement and could, at some point, cross paths. Aircraft entering the ground sector may have been rerouted due to the presence of the tractor. An exception was made for the Medium scenario with AF676SD, as this aircraft later became the cause of the taxiway closure event and thus was removed from consideration. On average,

On average, the tractors across both scenarios finished their routes with a normalized taxiing time of 1.19. Despite the shorter distance, the average normalized taxi time for SDH was 1.24, in comparison to 1.14 for HTR. However, the average normalized taxiing time of the aircraft in the ground sector was far longer. This effect may also be due to the change in configuration warning. The number of affected aircraft for each scenario was 5 and 6 for Medium and Hard, respectively. Figure 23 summarizes the average normalized taxi time for the tractors for all participants.

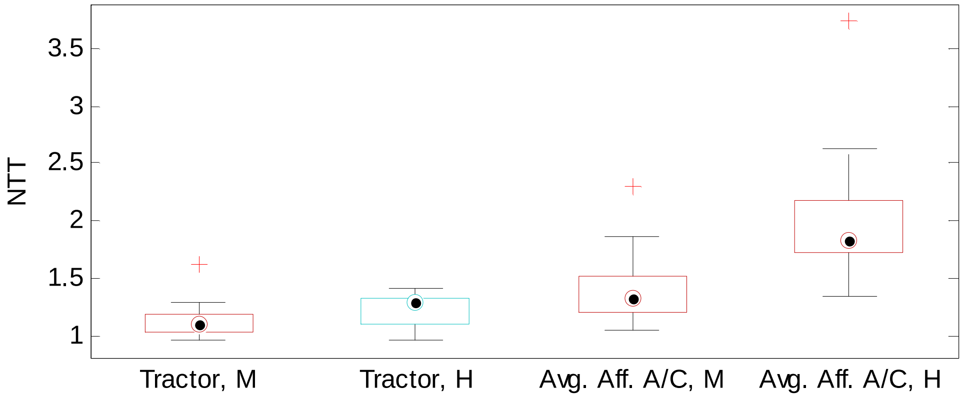

Figure 23: Normalized taxiing time for the tractors and affected aircraft (average).

4.9.1.5 Taxiway Closure

The taxiway closure event occurred in 16 out of 24 cases, but the taxiway closure in the Hard scenario turned out to be ineffective due to the lack of aircraft affected by this blockage. Nevertheless, a taxiway blockage does contribute to another dynamically unfolding event that the participant must monitor. Figure 24 illustrates the location of this closure and the other aircraft that were affected by it. In addition to monitoring this closure, the participant must reroute aircraft, potentially disturbing other routes.

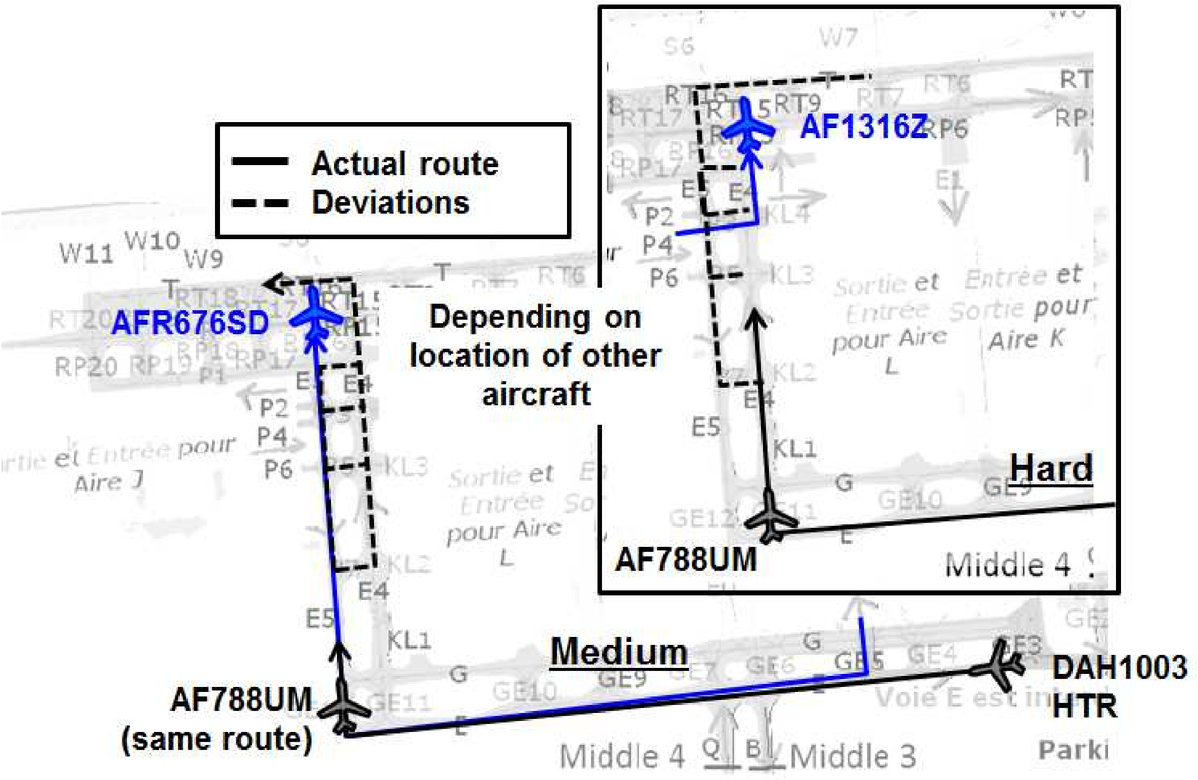

Figure 24: Visualization of the taxiway closure event.

The normalized taxiing time for the affected aircraft (AF788UM in both scenarios, AF758KG, DAH1003, HTR for Medium) were averaged to determine the impact. On average across both scenarios, the normalized taxi time was 1.74. Figure 25 illustrates the average normalized taxiing times for each participant.

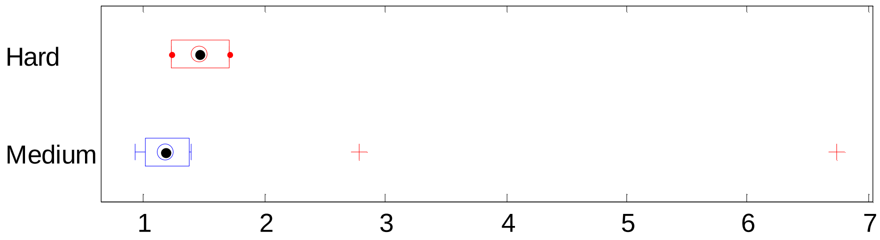

Figure 25: Normalized taxiing time of the affected aircraft in the taxiway closure event.

4.9.1.6 Change in configuration

The change in configuration differed between participants as their performances varied. However, such performance variation is hypothesized to occur in experiments 2 and 3 with the inclusion of new technologies. Therefore, to simplify analysis, all active aircraft between the moment of the QFU warning (T+5 minutes) and the start of the new configuration (T+15 minutes) were included in the analysis, for all participants, regardless of whether the aircraft was specifically rerouted due to the change in QFU. Participants were told that the change would be in effect in ten minutes from the moment of the alert. Figure 26 illustrates the paths of the affected aircraft (departures in blue, arrivals in red) during this time window.

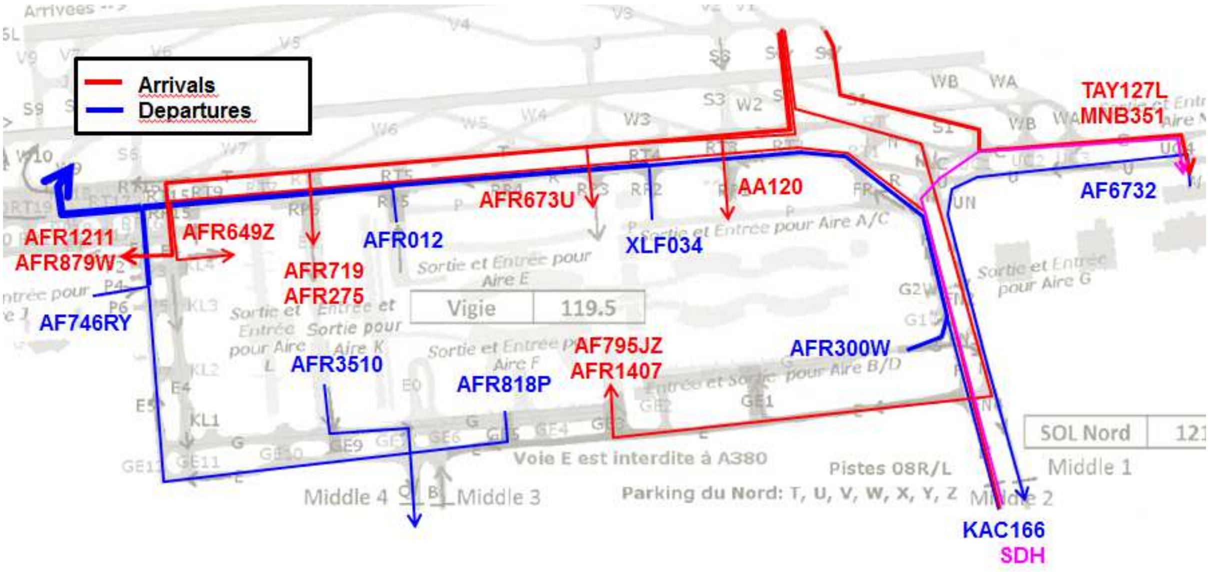

Figure 26: Visualization of the Change in configuration event.

In general, all participants continued sending aircraft on the old configuration (west) until the moment of the new configuration (east), except for one participant. This participant, once receiving the warning, began to prepare all departures for the east configuration, even delaying taxiing for several. This strategy minimized the taxiing time and the risk of rerouting aircraft headed towards the South. It is particularly effective in this scenario as a departure manager was not employed, meaning there were no specific timeslots that needed to be fulfilled.

The average normalized taxiing time for all participants was 2.19, with some participants having averages as high as 3.44. The lowest average normalized taxi times were 1.29 and 1.35. It is evident from the normalized taxi times of aircraft during this period that while most aircraft were well managed (even finding a shorter route and thus minimizing the taxiing time), there were several aircraft that needed to be rerouted, thus increasing the taxiing time. Figure 27 summarizes the average normalized taxiing time of the affected aircraft for each participant.

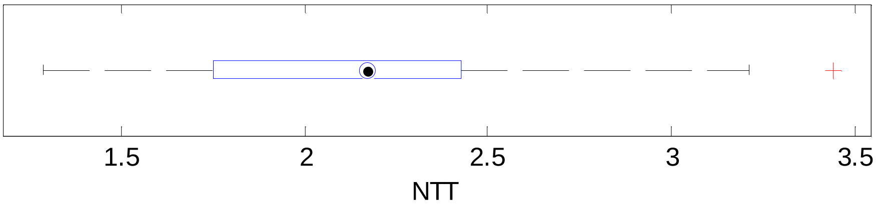

Figure 27: Average normalized Taxiing time for aircraft during the change in configuration event.

5 Discussion

The goal of this first experiment was to have an understanding of the current performance with today's ATC technology in France and to understand the shortcomings of this technology as to improve the design of MoTa. Overall, this experiment demonstrated that the two scenarios that we have chosen are sufficiently difficult. This range of activity allows us to better appreciate when and where MoTa provides the most significant contribution to the ATC task. The scenario complexity had a significant effect on self-reported workload, situation awareness, and the percentage of aircraft correctly treated. Additionally, the scenario complexity has a significant impact on the deviation from the ideal trajectory. There was not a significant effect on heart rate response, the normalized taxiing time, and the number of stop and gos. The lack of significance on the heart rate response is most likely due to the length of the scenario and the variations in the initial five minutes of each scenario. It is not possible to maintaining an elevated heart rate for durations as long as our scenarios.

The effect of scenario complexity must be spoken in regards to CDG experience, as knowledge and familiarity with the airport may decrease the amount of contributions offered by MoTa. Furthermore, the impact of the Hard scenario may also be softened due to experience. A known limitation of this study is the low number of ATCOS from CDG. As one of the secondary contributions of MoTa is to assist in the transition to newer technologies, we find that having this range of familiarity provides interesting inputs into the adaptability of this technology. When evaluating just the ATCOs with CDG experience, the scenario complexity only has a mildly significant effect on the percentage of aircraft correctly treated, workload, and situation awareness. While the effect is not as strong as desired, it is still there. Indeed, observations of participant performance also confirmed many of these statistical findings. The Hard scenario was deliberately designed to create more traffic than what is normally managed by a single person at CDG, an analogy for future traffic loads. Participants from CDG noted the excessive traffic; one even made a comment about needing to split the sector (as is commonly done in real operations). Participants not from CDG were even more overwhelmed, particularly at the moment of the configuration change.

Most participants were observed to be unable to sort the numerous reprinted strips while still addressing the traffic calls, and several chose to ignore the strips and work directly off of the radar screen. A few participants began to write notes on the back of the paper strips in lieu of sorting through those that had been printed. In addition to these technological difficulties, there were several common mistakes observed. Participants, on occasion, would mistakenly send aircraft to the wrong end of the airport and would be corrected by the pilot. The number of aircraft tickets made it difficult to locate the exact location on the radar map. The flight list of departing flights (Déport d'Information de Supervision et de Clairance pour les Utilisateurs dans les approcheS, DISCUS) was not observed to be used. The window was used at times by the participants from CDG, but infrequently by those not from CDG. Researchers were told specifically by at least three participants that they never looked up at the window. These interactions with the current level of technology will be taken into account during the development of the MoTa platform.

In addition to the information from paper technology and the out-the-window view, participants had trouble managing information on the screen. Similar to the radar maps, there were labels with each aircraft. Aircraft close together made it difficult to see individual pieces of information. Such problems will only increase with the addition of the taxibots (experiment 3). There are currently 10 such vehicles proposed to be added to the scenario.

Five operational events were introduced into the scenarios in order to explore more specific problems that are within the umbrella of the ground controller's responsibilities. The performance data for each of these events has been noted and will be compared to the equivalent scenarios in the next two experiments. The execution of the towed aircraft, the restricted zone, and change in configuration events were successfully deployed for almost all participants. However, the closed taxiway and the pilot error events were not present for all participants during the Hard scenario. As such, the frequency of events will likely be taken into account as a possible indicator of the effect of the interface and the automation, as the occurrence of these events was correlated with performance.

Experiments 2 and 3 will be run similarly to Experiment 1, with a few exceptions:

- Inclusion of additional material in the practice session to teach the participants the interface, and later, the taxibots
- Inclusion of Datalink-capable and Taxibot-compatible vehicles in the aircraft fleet: Current numbers propose a fleet composition of 85% datalink-equipped and 15% non-datalink. There are 10 taxibots proposed, with 10 aircraft using the taxibot system for the Medium scenario, and 20 aircraft for Hard.
- Change all callsigns of aircraft: This change will be implemented to avoid any confusion between researchers but also to eliminate any possible bias from repeat participants.
- Determine reasonable equivalents to the events in the first experiment: There are several considerations that must be taken into account. If the same participants are used, there is a possibility of residual association, thus an anticipation of the event itself. Additionally, it must be decided for the cause of certain events, such as the taxiway closure – does this event come from a temporary aircraft problem or a temporary taxibot problem, or perhaps another source?

It is likely that many of the same participants employed in experiment 1 will perform experiments 2 and 3, but the exact number is uncertain. Experiments 2 and 3 should start in October 2015, with the final report in January 2016.

6 Conclusion

The present baseline experiment, in addition to understanding the shortcomings of today's air traffic control technology in France, was also used to validate two scenarios, Medium and Hard. The experiment was conducted at the Ecole Nationale de l'Aviation Civile with 12 air traffic controller instructors with experience from airports around France. Five participants were from Charles-de-Gaulle, the airport that the simulation is based on. Each participant performed the two 35-minute ground taxiing scenarios. Both scenarios varied by the number of aircraft and different operational events (restricted zone, pilot errors, closed taxiway, change in configuration, towed aircraft). The scenario complexity and the simulator execution were both validated.

The results of this experiment demonstrate that the two scenarios are significantly different with respect to scenario complexity. Participants correctly treated a smaller percentage of aircraft in the Hard scenario compared to the Medium scenario. They reported having more workload and less situation awareness in Hard than Medium. The average taxiing experience (i.e. normalized deviation from the ideal trajectory) was significantly longer for aircraft in the Hard scenario than the Medium. While globally the average heart rate response was not found to be significant, several participants were measured to have more elevated heart rates during the Hard scenario as compared to the Medium. In summary, scenario complexity was determined to be significant on these variables:

- Percentage of aircraft
- Deviation from the ideal trajectory
- Situation awareness
- Workload

Five operational events were introduced in these scenarios to allow for further exploration of the effect of automation on the types of problems air traffic controllers must face in addition to the principle task of directing aircraft. The performances of these five events over the two scenarios have been recorded and will be compared to those measured in experiments 2 and 3. The primary comparison between the experiments with respect to these events is the delays in taxiing time for the affected aircraft (i.e. average normalized taxing time). Additionally, the global problem-solving of the participants will be compared.

Experiments 2 and 3, testing the effect of the interface and the taxibots respectively, will occur in October 2015.

7 References

Reference to main documentation, delete if not required

[1] Z. Chua, “E.02 24-MOTA-D3.1-Experimental Protocol”. 03 July 2015.

[2] Directorate General for Civil Aviation, Direction des Services de la navigation aérienne. “Manuel d’Exploitation TWR/APP”, DO/SNA-RP/CDG/SE. 2008-2009.

[3] Z. Chua, M. Causse, M. Cousy, and F. Andre. “Modulating Workload for Air Traffic Controllers during Airport Ground Operations”. HFES Annual Meeting 2015. Los Angeles, California, USA. 26-30 October 2015.

[4] Z. Chua, F. Andre, and M. Cousy. “Development of an ATC Tower Simulator to Simulate Ground Operations”. AIAA Aviation 2015. Modeling and Simulation Conference. Dallas, Texas, USA. 22-26 June 2015.

[5] S.G. Hart, S.E. Lowell, “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.” Advances in Psychology, Vol 52, 139-183. 1988.

[6] R.M. Taylor, “Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design.” AGARD, Situational Awareness in Aerospace Operations, 1990.

[7] D.M. Dehn, “Assessing the Impact of Automation on the Air Traffic Controller: The SHAPE Questionnaires”, Air Traffic Control Quarterly. Vol. 16, No. 2. 127-146. 2008

[8] M. Tarvainen, J. Niskanen. “Kubios HRV version 2.0 user’s guide”. Department of Physics, University of Kuopio, Kuopio, Finland. 2008.

Appendix A Title of the appendix

A Tables of Data used for analyses

B List of Aircraft and Characteristics for each scenario

Medium								
call SOL	**Callsign**	**Type**	**Type**	**Rwy**	**Parking**	**Dep**	**Arr**	**SID**
00:00:00	FIN8TR	Depart	A320	27L	D18	LFPG	EFHK	NURMO
00:00:15	EZY238X	Depart	A320	26R	B12	LFPG	EDDM	BUBLI
00:00:30	AFR599F	Depart	E170	26R	J16	LFPG	LJLJ	BUBLI
00:02:10	BAW307	Arrive	A319	26L	A04	EGLL	LFPG	MOPAR
00:03:00	KLM12P	Depart	B737	27L	F16	LFPG	EHAM	NURMO
00:05:40	DAL27	Arrive	B767	26L	**E26**	KATL	LFPG	MOPAR
00:06:10	AFR007	AFZONE	A380	27R	L71	LKPR	LFPG	LORNI
00:08:00	SWR67X	Arrive	A320	26L	A08	LSGG	LFPG	OKIPA
00:08:00	FIN874P	DERROR	A321	26R	D20	LFPG	EFHK	OPALE
00:09:50	AF689VC	Arrive	A318	26L	F10	LFMT	LFPG	BANOX
00:11:00	BEE493K	Depart	E170	27L	G30	LFPG	EGJJ	EVX
00:11:30	JAT240	Arrive	B737	26L	B14	LYBE	LFPG	OKIPA
00:12:00	AFR676SD	DPANNE	A320	26R	F02	LFPG	LFST	BUBLI
00:12:00	HTR	DTOWED	B743	27L	J10	LFPG	LFPG	OKIPA
00:13:00	AF651PQ	Arrive	A320	26L	F80	LFLL	LFPG	OKIPA
00:13:00	AF788UM	Depart	A319	26R	F80	LFPG	LFLL	ERIXU
00:14:10	AFR009	Arrive	B772	27R	K01	KJFK	LFPG	MOPAR
00:14:30	DAH1003	Depart	A330	26R	B18	LFPG	DAAG	LATRA
00:15:50	AFR6729	Arrive	B747	26L	M11	RKSI	LFPG	LORNI
00:18:00	AFR1710	Depart	A319	27L	F26	LFPG	EDDH	NURMO
00:18:30	AFL261	Arrive	A320	27R	C06	UUEE	LFPG	LORNI
00:20:30	AF758KG	Depart	E170	26R	**J37**	LFPG	LFRS	LGL
00:21:30	AFR139	Arrive	B772	27R	E14	HECA	LFPG	OKIPA
00:23:00	KLM38P	Depart	B737	27L	F04	LFPG	EHAM	NURMO
00:23:00	AFR981	Arrive	B772	26L	K21	RKSI	LFPG	LORNI
00:25:30	EZY808H	Depart	A319	26R	B03	LFPG	DTTA	ERIXU
00:25:50	FIN879Q	Arrive	A321	27R	**D19**	EFHK	LFPG	LORNI
00:26:40	AFR267	Arrive	B773	26L	K03	OMDB	LFPG	OKIPA
00:31:30	EZY343T	Arrive	A319	26L	D12	EDDM	LFPG	LORNI
00:33:00	AF645PQ	Arrive	A318	26L	F86	LFLL	LFPG	OKIPA
00:34:10	AFR3831	Arrive	B772	27R	K19	FIMP	LFPG	OKIPA

Hard								
call SOL	**Callsign**	**Type**	**ACType**	**Rwy**	**Parking**	**Dep**	**Arr**	**SID**
00:00:10	EZY465W	Depart	B747	26R	K01	*LFPG*	HECA	BUBLI
00:00:25	SWR633	Depart	A320	26R	W07	*LFPG*	LFLC	ERIXU
00:00:30	AAL41	Depart	B767	27L	E34	*LFPG*	KPHL	LGL
00:01:30	AFR300W	Depart	A320	26R	B03	*LFPG*	LPPR	LGL
00:01:50	AFR818P	Depart	A320	26R	F02	*LFPG*	LSZH	BUBLI
00:01:55	AFR012	Depart	B772	26R	E14	*LFPG*	KJFK	LGL
00:02:10	BAW308	Arrive	A319	26L	A12	EGGW	*LFPG*	MOPAR
00:02:20	AF737EW	Arrive	A318	26L	F30	DABB	*LFPG*	OKIPA
00:02:30	TAY127L	Arrive	B757	26L	M15	OMAA	*LFPG*	LORNI
00:02:30	SDH	DTOWED	B743	27L	M10	LCLK	*LFPG*	OKIPA

00:03:45	THY9JQ	Depart	A343	27L	K05	*LFPG*	KEWR	EVX
00:03:50	AF746RY	Depart	A321	26R	J18	*LFPG*	DTTA	ERIXU
00:04:00	AFR378	Depart	A330	26R	A02	*LFPG*	LWSK	LATRA
00:04:40	DAL229	Depart	B772	26R	D19	*LFPG*	DIAP	ERIXU
00:05:00	**QFU Warning**							
00:06:10	AAL120	Arrive	B757	26L	C03	KLAX	*LFPG*	MOPAR
00:06:30	AFR442Z	Depart	B747	27L	M08	*LFPG*	ZBAA	NURMO
00:06:35	KAC166	Depart	A318	26R	U08	*LFPG*	LIRQ	OKASI
00:07:40	AFR1211	AQFUWN	E170	26L	J33	EGBB	*LFPG*	MOPAR
00:09:10	AFR879W	Arrive	E190	26L	J10	LIPE	*LFPG*	OKIPA
00:09:10	UAE72	Depart	B773	26R	C02	*LFPG*	DBBB	ERIXU
00:09:20	AF788UM	Depart	A319	26R	F80	*LFPG*	OERK	BUBLI
00:10:15	ACA1901	Depart	A320	26R	U05	*LFPG*	LSGG	PILUL
00:10:40	AFR719	Arrive	B772	26L	K21	RKSI	*LFPG*	LORNI
00:11:05	FIN866L	Depart	E170	27L	G24	*LFPG*	LFRN	EVX
00:11:05	BEE743F	Depart	E170	27L	G24	*LFPG*	LFRN	EVX
00:12:10	AFR673U	Arrive	A321	26L	E10	LFML	*LFPG*	OKIPA
00:13:00	AFR980X	Depart	E190	27L	J02	*LFPG*	LKPR	RANUX
00:13:40	AF795JZ	AERROR	A319	26L	F04	DAOO	*LFPG*	BANOX
00:15:00	**Configuration E in effect**							
00:15:00	AFR418	Depart	B772	08L	E20	*LFPG*	*EDDH*	BUBLI
00:15:10	MNB351	ARRSTP	A342	26L	M07	KATL	*LFPG*	MOPAR
00:16:10	AFR649Z	AFZONE	A380	27R	L61	MMMX	*LFPG*	MOPAR
00:17:05	CSN348	Depart	A330	08L	Z03	*LFPG*	*LFLL*	ERIXU
00:17:10	AFR1316Z	DPANNE	CRJ7	08L	J35	*LFPG*	LJLJ	BUBLI
00:19:30	XLF034	Depart	A330	09R	L73	*LFPG*	*DAAG*	LATRA
00:19:50	AFR406	Depart	B772	08L	F03	*LFPG*	DTTA	ERIXU
00:22:45	AFR3510	Depart	B737	08L	B12	*LFPG*	*EHAM*	NURMO
00:22:55	AFR1407	Arrive	A319	27R	F34	LICJ	*LFPG*	OKIPA
00:23:30	LBT315	Depart	A320	08L	Q12	*LFPG*	*EFHK*	NURMO
00:23:30	AFR6732	Depart	B747	08L	M10	*LFPG*	ZBAA	NURMO
00:24:25	AFR275	Arrive	A320	27R	K53	OEJN	*LFPG*	OKIPA
00:25:00	AFR1886	Depart	E190	09R	J10	*LFPG*	OERK	BUBLI
00:25:55	AFR587	Arrive	A320	09L	F96	ESSA	*LFPG*	LORNI
00:26:30	EZY596K	Arrive	A319	08R	B06	LYBE	*LFPG*	OKIPA
00:27:15	AF678BF	Depart	A320	08L	J40	*LFPG*	LKPR	RANUX
00:28:55	EZY965V	Arrive	A319	09L	B05	EDDM	*LFPG*	LORNI
00:29:00	AFR231R	Arrive	CRJ7	08R	J35	DABB	*LFPG*	OKIPA
00:29:00	AFR6719	Depart	B767	09R	A10	*LFPG*	LFRS	LGL
00:30:30	BAW306	Arrive	A321	08R	A10	*EGLL*	*LFPG*	MOPAR
00:31:00	AFR933M	Depart	A320	09R	F32	*LFPG*	LFRS	LGL
00:32:00	AFR128U	Arrive	A320	08R	F16	*LICJ*	*LFPG*	OKIPA
00:34:45	BEE475U	AQFUNW	A321	09L	G14	*RKSI*	*LFPG*	LORNI

C Restricted Zone

The table below notes the NTT for the A380s presented to participant in each scenario. The box in gray indicates a recording error. In this situation, the event was performed, but there was an error with the data logging software.

	m: AFR007	**h: AFR649Z**
CARFL	0.99	0.99

DOVLO	1.13	1.54
GUSTO	1.00	0.58
MAIJU	1.12	1.06
MESCH	1.09	■
PANBR	1.04	1.03
PASLI	0.98	*
PIEAN	0.98	2.02
RASFR	▨	1.16
SAIJU	0.99	0.99
SANPH	1.03	1.65
ZIMOL	0.86	0.29
Average	1.02	1.13
■	Event did not occur	
▨	Data recording error	
*	Used later for Taxiway closure	

D Pilot Error

The table below notes the reaction and resolution times for each participant during the pilot error event.

	Medium (mins)		**Hard (mins)**	
	Reaction	**Resolution**	**Reaction**	**Resolution**
CARFL	00:24	00:12	■	■
DOVLO	00:43	00:08	01:49	00:09
GUSTO	00:44	00:50	00:26	00:56
MAIJU	00:15	00:40	▨	▨
MESCH	▨	▨	■	■
PANBR	00:23	00:18	■	■
PASLI	00:17	00:32	00:01	00:16
PIEAN	01:09	00:40	00:48	00:19
RASFR	00:25	01:21	■	■
SAIJU	00:18	00:13	01:56	00:15
SANPH	00:30	02:23	■	■
ZIMOL	00:46	00:46	▨	▨
Average	**00:32**	**00:44**	**01:00**	**00:23**
■	Event did not occur			
▨	Data recording error			

E Towed Aircraft

The table below notes the NTT for the tractors and aircraft on the ground at the time of the tractor.

	CARFL	DOVLO	GUSTO	MAIJU	MESCH	PANBR	PASLI	PIEAN	RASFR	SAIJU	SANPH	ZIMOL
m: HTR	**1.01**	**1.17**	**1.10**	**1.03**	**1.62**	**1.04**	**1.19**	**0.96**	**1.11**	**1.01**	**1.29**	**1.19**
AF689VC	1.71	1.85	2.27	1.66	1.78	1.54	1.80	1.35	0.41	1.47	2.89	1.81
JAT240	1.62	1.57	1.48	1.07	1.96	1.35	1.28	1.62	1.43	1.22	1.44	1.22
AF788UM	0.98	1.04	2.71	1.26	1.41	0.92	1.11	1.06	1.30	6.74	1.72	1.15
DAH1003	1.03	1.14	1.22	1.43	1.08	0.14	1.16	1.07	1.04	1.01	0.14	0.96
AFR1710	1.24	1.06	1.60	1.03	1.38	1.26	1.26	0.85	1.00	0.99	1.43	0.95
Avg. Affected A/C	1.32	1.33	1.86	1.29	1.52	1.04	1.32	1.19	1.04	2.29	1.52	1.22
h: SDH	**1.07**	**1.31**	**1.33**	**1.28**	**1.26**	**1.09**	**1.31**	**1.41**	**1.41**	**1.10**	**0.95**	**1.33**
AF737EW	1.24	1.44	1.01	1.30	■	1.01	■	1.03	1.06	1.38	1.12	■
SWR633	3.84	3.26	3.64	3.51	3.11	3.54	3.19	2.85	0.97	3.28	2.87	1.42
AFR300W	1.53	1.41	1.16	1.18	1.89	2.01	1.39	2.67	1.84	1.07	1.64	1.08
TAY127L	1.07	1.31	1.33	1.28	1.26	1.09	1.31	1.41	1.41	1.10	0.95	1.33
FIN866L	1.35	1.18	■	2.83	0.32	■	1.23	5.17	1.33	1.54	■	0.99
AFR6732	2.07	1.44	11.63	3.14	2.39	3.07	1.78		1.42	2.20	2.90	2.12
Avg. Affected A/C	1.85	1.68	3.75	2.21	1.79	2.15	1.78	2.63	1.34	1.76	1.89	1.39

F Taxiway closure

The table below notes the NTT for the taxiway closure event.

	Medium	Hard
CARFL	1.00	■
DOVLO	1.26	■
GUSTO	2.78	■
MAIJU	1.39	*
MESCH		■
PANBR	6.74	■
PASLI	1.18	1.70
PIEAN	1.06	■
RASFR	1.30	■
SAIJU	0.98	■
SANPH	0.93	■
ZIMOL	1.10	1.22
Average	1.79	1.46

■ Event did not occur

* occurred, but no effect

G Change in Configuration

The table below notes the NTT for all of the affected aircraft during the change in configuration event.

	CARFL	DOVLO	GUSTO	MAIJU	MESCH	PANBRI	PASLI	PIEAN	RASFR	SAIJU	SANPH	ZIMOL
AFR300W	1.53	1.41	1.16	1.18	1.89	2.01	1.39	2.67	1.84	1.07	1.64	1.08
AFR818P	3.36	1.77	5.36	1.61	2.13	0.94	1.54	0.99	1.18	1.45	1.24	1.07
SDH	1.07	1.31	1.33	1.28	1.26	1.09	1.31	1.41	1.41	1.10	0.95	1.33
TAY127L	1.64	1.50	2.01	1.16	1.93	1.06	1.11	1.47	1.13	1.03	1.55	0.34
AFR012	1.12	1.13	2.08	5.64	6.18	[redacted]	1.38	2.13	1.05	1.60	1.92	1.59
AF746RY	12.11	5.89	[redacted]	7.00	3.08	[redacted]	3.69	0.49	2.03	3.76	[redacted]	1.00
XLF034	[redacted]	1.04	1.24	1.11	1.05	4.29	2.51	2.05	2.43	2.61	2.55	1.33
AAL120	0.95	0.84	0.80	1.29	0.89	0.85	0.89	1.00	[redacted]	0.84	2.07	0.91
AFR3510	4.81	2.69	19.38	2.81	25.44	[redacted]	2.50	2.94	0.94	14.75	2.81	1.00
AFR1211	1.19	1.91	1.00	1.27	1.52	1.00	0.99	1.27	1.61	0.98	3.43	1.02
AFR6732	2.07	1.44	11.63	3.14	2.39	3.07	1.78	[redacted]	1.42	2.20	2.90	2.12
KAC166	7.50	5.75	[redacted]	6.76	6.66	1.23	8.70	5.41	1.17	8.33	8.61	3.85
AFR879W	1.00	1.45	0.97	1.88	1.60	0.99	2.19	2.34	1.36	0.99	1.15	2.23
AFR719	1.02	1.40	1.00	1.82	1.33	1.00	1.87	3.94	1.01	1.00	1.53	1.00
AFR649Z	0.99	1.54	0.58	1.06	[redacted]	0.99	4.72	2.02	1.16	1.03	1.65	[redacted]
AFR673U	0.84	0.79	0.80	1.78	4.80	3.53	2.75	1.33	1.11	0.79	0.79	0.79
UAE72	2.42	3.37	[redacted]	4.22	[redacted]	[redacted]	1.37	0.99	0.93	1.60	[redacted]	2.72
AF795JZ	0.99	1.43	0.97	1.04	1.28	0.54	1.49	1.10	0.98	1.22	0.84	1.29
AFR980X	1.17	1.12	[redacted]	1.17	[redacted]	[redacted]	1.14	1.15	1.18	[redacted]	[redacted]	1.02
MNB351	1.47	1.12	0.34	2.05	0.34	[redacted]	1.02	1.13	1.19	1.03	1.06	0.96
AFR275	1.01	0.48	4.79		0.42	[redacted]	1.75	1.60	1.17	1.16	[redacted]	0.99
AFR1407	2.00	3.03	2.44	2.31	1.10	[redacted]	1.85	2.19	0.80	1.69	[redacted]	0.78
Average	**2.39**	**1.93**	**3.21**	**2.46**	**3.44**	**1.61**	**2.18**	**1.89**	**1.29**	**2.39**	**2.16**	**1.35**

-END OF DOCUMENT-