Task measures for air traffic display operations

Tan, Shi Yin; Chen, Chun-Hsien; Lye, Sun Woh; Li, Fan

2020

Tan, S. Y., Chen, C.-H., Lye, S. W., & Li, F. (2020). Task measures for air traffic display operations. *Proceedings of International Conference on Intelligent Human Systems Integration*, 903, 183-189. doi:10.1007/978-3-030-39512-4_29

<https://hdl.handle.net/10356/146620>

https://doi.org/10.1007/978-3-030-39512-4_29

© 2019 Springer Nature Switzerland AG. All rights reserved. This paper was published in 2020 IHSI International Conference on Intelligent Human Systems Integration and is made available with permission of Springer Nature Switzerland AG.

Downloaded on 26 Apr 2023 21:49:19 SGT

Task Measures for Air Traffic Display Operations

Shi Yin Tan^{1,*}, Chun Hsien Chen¹, Sun Woh Lye¹, Fan Li^{2*}

¹ School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Ave, Block N3, Singapore 639798

² Fraunhofer IDM Centre, Nanyang Technological University, 50 Nanyang Ave, Block NS1 Singapore 639798

Abstract. With rising growth in air traffic globally, advanced technologies are being developed to aid ATCOs in the managing and control of a foreseeable denser airspace. The need to perform holding stack management, a potential challenge to ATCO, especially during heavy traffic congestion owing to weather and runway conditions is expected to be more frequent. To mitigate this challenge, the use of 3D displays was suggested. This paper examines the performance impacts resulting from the adoption of 3D instead of 2D radar displays with regards to visual search and relative vertical positioning identification. Observations relating perceived increased in stress and workload by the participants are also made.

Keywords: Human Performance · Task Measures · Stress and Workload · Perspective Changes in Dimensional Representation

1 Introduction

Affordable commercial air travel has sparked rapid growth in the aviation industry in recent years [1]. Managing and controlling air traffic effectively and efficiently is becoming more important than ever before. Given that the primary role of an air traffic controller (ATCO) is to guide and direct air traffic flow, ATCO is presented with a vast amount of flight data daily. From these data, ATCO would then process them and make appropriate decisions. This is because their workload is dependent on the number of flights within their jurisdiction [2]. Thus, several advanced technologies have been implemented to reduce the workload of ATCOs, such as conflict resolution advisory aids [3], 3D radar display [4], and adaptive fatigue management [5]. Nevertheless, implementing these novel advanced technologies would substantially change the interactions between human and systems. In other words, the tasks and operational procedures of ATCOs will change with the introduction of advanced technologies.

In a conventional ATC setup, screens are made to display a plan view of the airspace with vertical separation of aircrafts denoted by altitude values next to the aircraft markers. In recent years, there have been various studies on the use of 3D radar displays that suggest evidence that a change in perspective can help to improve the ATCOs spatial awareness of traffic and terrain thereby better navigate aircrafts in the vertical plane [6]. Interface congestion or cluttering is also a problem with 2D radar displays when operations such as holding stack management is utilized [7]. In holding stack management, aircrafts are tasked to fly in a preassigned route while waiting for clearance to land. This preassigned route normally takes the form of an oval circuit with multiple aircrafts following the same route separated vertically. The close horizontal proximity on the screen of these aircrafts while carrying out a holding procedure poses problems to the ATCO viewing on 2D radar displays due to the overlapping of flight labels and markers.

It has been suggested that the use of 3D radar displays can help to reduce the cognitive loads of the ATCOs via the reduction of interpretative task as better and clearer mental picture can be derived from such displays. This is especially so in managing congested airspace or in holding stack management [6]. Furthermore, the alleviation of congestion of markers in the display using 3D radar displays can help ATCOs make better judgements on vertical arrangements. However, little is reported on the effects of 3D displays on ATCO's acceptance and performance. This paper seeks to investigate into performance measures relating to accuracy and task completion time between 2D and 3D display interfaces in an ATC setting.

2 Methodology

2.1 Experimental Design

A laboratory-based experiment composing of 16 university-level student subjects was conducted to study into the two measures of accuracy and completion times when performing various scenario tasks using 2D and 3D displays. Each exercise consists of a set of eight scenarios. On accuracy, this is based on the number of correct answers provided by the participants when queried. For task completion time, this is based on the recorded timestamps that captures the start of task till the final mouse click that signals the end of the exercise. Besides this, NASA-TLX questionnaires were also used where the participants are required to complete the study based on a Likert-type rating scale of (0-9) aimed to measure their perceived mental workload and stress when undertaking the various scenario tasks.

The experiment was conducted using a real-time simulator, the NLR ATM Research Simulator (NARSIM), to generate and simulate air traffic scenarios in real time. The simulated scenarios were displayed on two display monitors. Scenarios used in this study are based on holding stack management, a scenario which occurs when multiple aircrafts upon ATCO's instructions are required to fly in a holding pattern at separate altitudes. 2D and 3D visual interfaces are used in this experiment as shown in Fig. 1.

Fig. 1. 2D Visual Orientation (Left), 3D Visual Orientation (Right)

2.2 Experimental Procedure

Fig. 2 shows the experimental flow where participants would first be required to undergo several calibration procedures in relation to the EEG headset and eye tracker that would be used in this exercise. This is then followed by a training period in which participants familiarize with the experimental environment by running through several simulated scenarios.

Fig. 2. Experimental Flow of Exercise to be Conducted

Once familiar, the actual experiment exercise commences which comprises 8 test scenarios of varying intensity and dimensional information representation as highlighted in Fig. 3.

Scenario	First Screen	Second Screen	Number of Dots
1	2D orientation	2D orientation	3 points
2	3D orientation	2D orientation	3 points
3	3D orientation	3D orientation	3 points
4	2D orientation	3D orientation	3 points
5	2D orientation	2D orientation	4 points
6	3D orientation	2D orientation	4 points
7	3D orientation	3D orientation	4 points
8	2D orientation	3D orientation	4 points

Fig. 3. A set of Eight Scenarios to be Performed

The procedure within each scenario consists of two phases as follows:

- (a) **Memory Phase:** Participants were first asked to memorize key flight details such as the callsigns and altitude of the aircraft markers that are presented in the left display for 10 seconds before the screen will be turned off. For example in Scenario 1, the left display would present a particular airspace with three dots denoting the flights with its callsigns and flight levels highlighted in 2D orientation.
Test Phase: The right display is then activated, displaying the same airspace and flight activity and information. Participants are then required to carry out two tasks namely:
T1: A visual search task where one is to locate the aircraft markers in alphabetical order based on their callsigns.
T2: Relative Position Identification when one is to select an aircraft at a higher altitude out of 2 highlighted flight markers.
- (b) The participant is required to give his/her answer for T1 and T2. Each (Memory and Test) phase is repeated 15 times per scenario per participant.
- (c) At the end of the scenario, the participant is required to fill up a set of NASA-TLX questionnaires used to measure the perceived mental workload and stress for that scenario.
- (d) A 5 minutes break would then be instituted before the commencement of the next scenario. This process repeats itself until a participant completes all the eight scenarios.

It is to be noted that in the first 4 scenarios, different combinations of 2D and 3D orientation representation were adopted. For scenarios 5-8, they are similar to scenarios 1-4 except that it functions at a higher intensity level where the number of aircraft markers are increased from 3 to 4. These intensity values were chosen as these values represent the limits of spatial working memory of young healthy adults. It was reported that a young adult is found to be only capable of maintaining 3 to 4 simple visual objects in working memory [8][9].

3 Results & Discussion

3.1 Response Accuracy

Accuracy in response was based on the answers provided by the participant for the 2 test tasks (T1 and T2) in each of the scenarios. For each test task, 240 responses were derived from all the participants. Fig. 4 shows the mean or average percentage of correct responses over the total number of participants' responses for each test task in the various scenarios. To evaluate situations having low and high markers, scenarios 1 and 5 of 2D>2D are used as reference. 2-tailed paired t-test was then conducted between these references with the other 3 display combinations (3D>2D, 3D>3D and 2D>3D) to find the significance of the values. A p score < 0.05 is deemed to be significant whereas score > 0.05 is regarded as null.

Intensity Level		Low (3 Markers)				High (4 Markers)			
Scenario/Session No.		1	2	3	4	5	6	7	8
Orientation of Display (L>R)		2D>2D	3D>2D	3D>3D	2D>3D	2D>2D	3D>2D	3D>3D	2D>3D
Task 1	Mean	0.9960	0.9170	0.9875	0.9292	0.9460	0.8420	0.9790	0.8920
	Two-tailed P-value		0.0205	0.3332	0.0998		0.0007	0.1347	0.1544
Task 2	Mean	0.9330	0.9500	1.0000	1.0000	0.8920	0.8130	0.9750	0.9920
	Two-tailed P-value		0.4506	0.0009	0.0009		0.0318	0.0128	0.0012

Fig. 4. Average Task Accuracy

In visual search tasks (T1), t-test results showed that there were significant differences in accuracy registered for 3D>2D scenarios (2 & 6) with respect to 2D>2D scenarios ($p=0.0205$ and $p=0.0007$ for the low (3 dots/markers) and high intensity (4 dots/markers) levels respectively. Based on the mean of Scenario 2 (0.917) and Scenario 1 (0.996), there is a significant drop in accuracy in performing visual search tasks. This is also seen at higher intensity levels where lower accuracy and 0.842 for scenario 6 were registered compared to of 0.946 for scenario 5. One possible reason is that in scenario 2 and 6, the change in visual projection may have forced the participants to reorient the memorized airspace information which may be a factor for the decrease in accuracy. Furthermore, participants performed the worse in terms of accuracy when the 3D orientation is used for the memory phase. This coincides with a study that found significant detrimental effects of using 3D orientation whilst performing visual search tasks [10].

In identifying vertical relative positioning tasks (T2), all other scenarios register significant difference except for 3D>2D at low intensity level ($p=4.506$). Accuracy for T2 was higher in scenarios where the dimensional representation is either upgraded from 2D>3D or at 3D>3D orientation (Scenarios 3,4,7,8). It is observed that introduction of a depth cue in the 3D orientation allowed participants to visualize altitude separation of aircraft better thereby greatly enhanced performance in T2. Participants engaged in Scenario 6 (3D>2D) fared the worst on average. This would be due to a combination of perspective reorientation and engagement of a more complicated 3D display in the memory phase.

3.2 Task Completion Time

On task completion time, the measure is from the start of the test stage to the final click recorded in the test stage. The timings were recorded in milliseconds. Like the previous exercise on response accuracy, average values were obtained over 240 responses across all participants for T1 and T2 tasks. and represented in Fig. 5. These

average values are based on times where the participants have completed their tasks correctly.

Intensity Level		Low (3 Markers)				High (4 Markers)			
Scenario/Session No.		1	2	3	4	5	6	7	8
Orientation of Display (L>R)		2D>2D	3D>2D	3D>3D	2D>3D	2D>2D	3D>2D	3D>3D	2D>3D
Task 1	Mean	2011.658	2303.478	1997.386	2781.289	2258.968	2434.354	2353.641	3693.824
	Two-tailed P-value		0.0048	0.7638	0.0005		0.0038	0.0680	0.0000
Task 2	Mean	1495.254	1588.345	1268.742	1265.500	1905.397	2151.981	1399.420	1347.415
	Two-tailed P-value		0.3468	0.0396	0.0245		0.0370	0.0020	0.0014

Fig. 5. Average Completion Time

In visual search tasks (T1), values for 3D>3D scenarios (scenarios 3 and 7) were found to be insignificant ($p=0.7638$ and $p=0.068$) This could be owing to the single orientation setup in these scenarios. For scenarios involving dual orientation displays (Scenarios 2,4,6,8), participants completed the tasks on average slower than a single orientation setup for both low and high intensity levels.

For vertical relative positioning identification (T2), values for all other scenarios are found to be significant except for Scenario 2 ($p=0.3468$). Scenarios which utilized the 3D orientations for the test phase (3,4,7,8) registered faster completion time on average as 3D orientation display provides a better representation in terms of relative vertical positioning. However, in scenario 6, a longer completion time is observed when compared to scenario 5 ($2151.981>1905.397$). This may be attributed to the higher complexities of the 3D representation during the memory phase.

3.3 Self-Task Evaluation

For the self-evaluation section by the participants, NASA-TLX questionnaires were used. Each participant is required to evaluate the perceived stress and workload at the end of each scenario. Mean scores across all participants for each scenario along with the respective 2-tailed t-test p-values relative to the 2D>2D scenarios for each respective intensity level are shown in Fig. 6.

Intensity Level		Low (3 Markers)				High (4 Markers)			
Scenario/Session No.		1	2	3	4	5	6	7	8
Orientation of Display (L>R)		2D>2D	3D>2D	3D>3D	2D>3D	2D>2D	3D>2D	3D>3D	2D>3D
Stress	Mean	3.72	4.17	3.50	3.67	4.06	4.56	4.00	4.22
	P-Value		0.0278	0.1631	0.7903		0.0032	0.7903	0.2687
Workload	Mean	3.94	4.72	3.44	3.78	4.78	5.33	4.00	4.44
	P-Value		0.0016	0.0702	0.5921		0.0013	0.0302	0.1376

Fig. 6. Participants' Questionnaire Scores on Perceived Stress and Workload

From Figure 6, a significant higher mean is registered for scenarios 2 and 6 when compared to 2D>2D scenarios indicating that a higher level of stress was perceived by the participants. Noticeable p scores of $p=0.0278$ and $p=0.0032$ for scenarios 2 and 6 respectively were also registered which indicated a significant difference between the two compared scenarios. On workload, the mean scores of the participants show significantly difference for scenarios 2, (6,7) when compared with scenarios 1 and (5) respectively. This is also reflected in the p scores (< 0.05) for these scenarios.

When the increase in perceived stress and workloads in scenarios 2 and 6 are compared with response accuracy, a lower accuracy and slower completion time were observed in visual search tasks. For relative vertical positioning identification, significant increment in task completion times were also observed. One could postulate that higher perceived stress and workload can result in lower response accuracy and longer task completion times.

4 Conclusion

Visual search and relative vertical positioning tasks were conducted to study the performance effects on 2D and 3D displays. It was found that response accuracy and completion time could be used to identify the difficulties of various scenarios as experienced by the participants.

Significant differences were observed in terms of lower response accuracy and longer task completion time when using dual displays from 3D to 2D dimensional representations. On relative vertical positioning tasks, upgrade of dimensional representation from 2D to 3D displays or use of 3D display can help to improve accuracy and completion time due to the reduction of interface cluttering.

The reduction in response accuracy and speed of completion could be used as measures to gauge the stress and workload experienced by the subjects.

References

1. Release, I.A.T.A.J.P., IATA forecasts passenger demand to double over 20 Years. (2016).
2. Endsley, M., Rodgers, M.: Situation Awareness Information Requirements Analysis for En Route Air Traffic Control. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 38, 71-75 (1994).
3. Trapsilawati, F., Wickens, C., Qu, X., Chen, C.: Benefits of Imperfect Conflict Resolution Advisory Aids for Future Air Traffic Control. Human Factors: The Journal of the Human Factors and Ergonomics Society. 58, 1007-1019 (2016).
4. Liu, Y., Trapsilawati, F., Lan, Z., Sourina, O., Johan, H., Li, F., Chen, C., Mueller-Wittig, W.: Human Factors Evaluation of ATC Operational Procedures in Relation to Use of 3D Display. Advances in Human Factors of Transportation. 715-726 (2019).
5. Li, F., Xu, G. Li, F., Chen, C. H., Khoo, L. P.: Contextual information-based human fatigue prediction for integrated traffic control. Proceedings of International Conference on Computers and Industrial Engineering. (2018).
6. Wong, B., Rozzi, S., Boccalatte, A., Gaukrodger, S., Amaldi, P., Fields, B., Loomes, M., Martin, P.: 3D-in-2D Displays for ATC. 6th EUROCONTROL Innovative Research Workshop & Exhibition. pp. 47-62. EUROCONTROL Experimental Centre, Bretigny (2007)
7. Masotti, N., Persiani, F.: On the history and prospects of three-dimensional human-computer interfaces for the provision of air traffic control services. CEAS Aeronaut. J. 7(2), 149-166 (2016)
8. Cowan, N.: The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences. 24, 87-114 (2001).
9. Fukuda, K., Awh, E., Vogel, E.: Discrete capacity limits in visual working memory. Current Opinion in Neurobiology. 20, 177-182 (2010).
10. Popelka, S., Doležalová, J.: Non-photorealistic 3D Visualization in City Maps: An Eye-Tracking Study. Lecture Notes in Geoinformation and Cartography. 357-367 (2014).