Human factors evaluation of conflict resolution aid in future air traffic control

Fitri Trapsilawati

2017

Fitri Trapsilawati. (2017). Human factors evaluation of conflict resolution aid in future air traffic control. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/71503>

<https://doi.org/10.32657/10356/71503>

HUMAN FACTORS EVALUATIONS OF CONFLICT RESOLUTION AID IN FUTURE AIR TRAFFIC CONTROL

FITRI TRAPSILAWATI

School of Mechanical and Aerospace Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2017

Acknowledgements

Pursuing a PhD is a long and challenging process and it is absolutely not done solo. First of all, I would like to fully express my sincere gratitude for my advisors, Professors Chen Chun-Hsien, Christopher D. Wickens, Qu Xingda, and Park Taezoon. They wholeheartedly guided, supported, and encouraged me along my PhD journey.

I would also like to extend my gratitude to the Air Traffic Management Research Institute and the Civil Aviation Authority of Singapore for providing the experiment-related matters including the air traffic controller participants, the air traffic control simulator, and the research funding. I also thank Mr Erik Jan-Hartlieb from the National Aerospace Laboratory, Netherland, for providing the excellent technical assistance. Further, I am thankful to my friends at Nanyang Technological University for the stimulating discussions as well as the fruitful and unforgettable moments.

Last but not least, I am truly indebted to my family, especially my beloved mother and my awesome late father for their endless love and courage, as well as my kind parents-in-law for their cares. A special heartfelt gratitude goes to my wonderful husband who always cares and supports me.

Table of Contents

Acknowledgements	II
Table of Contents	III
List of Figures.....	VII
List of Tables	IX
List of Abbreviations	X
Summary.....	XIV
Chapter 1	1
Introduction.....	1
1.1 Research Background.....	1
1.1.1 Air traffic control.....	1
1.1.2 Automation in air traffic control.....	2
1.1.3 Automated conflict resolution	3
1.2 Research Objectives and Hypotheses.....	6
1.3 Research Scope	11
1.4 Dissertation Organization.....	11
Chapter 2	13
Literature Review	13
2.1 Air Traffic Control (ATC)	13
2.1.1 The Overview of ATC.....	13
2.1.2 Air Traffic Conflict.....	16
2.2 Human-Automation Interaction (HAI) in ATC	17
2.2.1 The Overview of HAI.....	17
2.2.2 Human Factors Concepts in HAI	20

2.2.3 Drivers of HAI in ATC.....	41
2.3 Conflict Detection and Conflict Resolution System	68
2.3.1 Support Tools for Conflict Detection and Conflict Resolution.....	68
2.3.2 Operational ATC Concepts for Conflict Resolution in ATC Facilities.....	72
2.4. Limitations in Studies on Conflict Resolution in ATC	86
Chapter 3	90
Research Framework.....	90
Chapter 4	99
Study 1: Effects of Conflict Resolution Aid Reliability and Time Pressure on Human Performance Variables of Air Traffic Controllers	99
4.1 Methods.....	100
4.1.1 Participants	100
4.1.2 Apparatus.....	100
4.1.3 Experimental design and procedure	104
4.1.4 Dependent Measures.....	109
4.1.5 Statistical Analysis	109
4.2. Results.....	110
4.2.1 Task performance measures	110
4.2.2 Subjective Ratings of Mental Workload and Trust	113
4.2.3 Aid Utilization Rate.....	113
4.2.4 Situation Awareness	113
4.3 Discussion	115
Chapter 5	120
Study 2: Interaction of Conflict Resolution Aid Reliability and Traffic Density in Future Air Traffic Control.....	120
5.1 Methods.....	120

5.1.1 Participants	120
5.1.2 Apparatus.....	121
5.1.3 Design.....	121
5.1.4 Tasks.....	122
5.1.5 Automation Implementation.....	124
5.1.6 Statistical analysis	126
5.2 Results	128
5.2.1 Task performance measures	128
5.2.2 Mental workload vs Subjective ratings of mental workload	130
5.2.3 Situation awareness	130
5.2.4 Subjective ratings of trust	131
5.2.5 Aid utilization rate	131
5.2.6 First failure effect	131
5.3. Discussion	133
Chapter 6	138
Study 3: Integration of a Vertical Situation Display with the Conflict Resolution Aid and their Interaction Effects on ATCO Performance	138
6.1 Methods.....	139
6.1.1 Participants	139
6.1.2 Apparatus.....	139
6.1.3 Experiment Design and Procedure	147
6.1.4 Dependent Measures.....	149
6.1.5 Analysis	150
6.2 Results	150
6.2.1 Task performance measures	150

6.2.2 Mental Workload	152
6.2.3 Situation Awareness	153
6.2.4 First failure effects.....	155
6.3 Discussion	156
Chapter 7	161
Conclusions.....	161
7.1. Key Findings and Contributions	161
7.2. Lessons Learned Across the Three Studies.....	166
7.3. Limitations and Future Directions.....	171

List of Figures

Figure 1. 1 Structure of Dissertation.....	12
Figure 2. 1 Air Traffic Control Communication Path (Durso & Manning, 2008)	14
Figure 2. 2 Air Traffic Separation	16
Figure 2. 3 Stages of Automation	22
Figure 2. 4 Trust Calibration Function (Lee and See, 2004)	26
Figure 2. 5 Open-Loop Model of ATCOs' Mental Workload (Loft, et al. 2007)	38
Figure 2. 6 Full-mode VSD (Prevot and Plamer, 2000)	56
Figure 2. 7 Co-planar VSD (Prevot and Palmer, 2000).....	57
Figure 2. 8 Electronic Vertical situation display (EVSD) (Vakil, et al, 1996).....	59
Figure 2. 9 Schematic Representation of Conflict Geometry	66
Figure 2. 10 Solving Conflict Process (Kallus, et al. 1999)	73
Figure 2. 11 Example of Resolution Trajectory (Erzberger, 2006)	79
Figure 3. 1. Research Framework	90
Figure 3. 2 Example of a SA probe	96
Figure 3. 3. Research model for experiment 3	98
Figure 4. 1 Primary radar display	101
Figure 4. 2 CRA display	103
Figure 4. 3 Example of a secondary conflict situation.....	108
Figure 4. 4 Conflict resolution performance (error bars indicate +/-1 SE).....	111
Figure 4. 5 Percentage of correct SA response (error bars indicate +/- 1 SE).....	114
Figure 5. 1 UAH condition	125
Figure 5. 2 UAL condition.....	125

Figure 5. 3 Conflict resolution performance across traffic densities (error bars indicate +/- 1 SE).....	129
Figure 6. 1 Primary Radar Display. Blue and green lines represent SID and STAR, respectively. Parallel yellow lines indicate extended runways for interception to the ILS localizer.....	141
Figure 6. 2 Flight Data Display. Yellow and blue rows indicate FPS for arriving and departing aircraft respectively.....	142
Figure 6. 3 The Plan View (a) and Vertical situation display (b) with the application of the HF principles.....	146
Figure 6. 4 Example of Conflict in the Unreliable Condition.....	148
Figure 6. 5 Conflict Resolution Performance (error bars indicate +/-1 SE)	151
Figure 6. 6. Interaction effect between automation conditions and display on SA probes accuracy (error bars indicate +/- 1 SE)	154
Figure 7. 1 Forrest plot analysis for performance (overall effect size of the automation condition across the three studies is shown).....	168
Figure 7. 2 Forrest plot analysis for workload (overall effect size of the automation condition across the three studies is shown)	169
Figure 7. 3 Forrest plot analysis for SA (overall effect size of the automation condition across the three studies is shown)	170

List of Tables

Table 2. 1 Levels of Automation of Decision and Action Selection (Parasuraman, Sheridan, Wickens, 2000).....	21
Table 2. 2 LOA Taxonomy for Human-Automation Performance in Dynamic, Multitask Scenarios (Endsley and Kaber, 1999).....	22
Table 2. 3. Relevant studies of VSD.....	55
Table 2. 4. Automation supports for conflict detection and resolution	69
Table 2. 5. Studies related to automation conflict resolution	76
Table 4. 1 The order of experiment conditions.....	104
Table 4. 2 Dependent measures under the various CRA conditions: Mean (SE).....	112
Table 5. 1 Experimental Design.....	122
Table 5. 2 Counterbalancing of the Experiment Conditions	128
Table 5. 3 Mean Probe Response Latency during the First Failure (Standard error in parenthesis)	133
Table 5. 4 The Summary of the Hypotheses Result	133
Table 6. 1 Mean Percentage of Resolved Conflict (SE in parentheses)	151
Table 6. 2 Mean Conflict Resolution Time (SE in parentheses)	152
Table 6. 3 Mean Workload Rating (SE in parentheses)	152
Table 6. 4 Mean Percentage of Timeouts (SE in parentheses)	153
Table 6. 5 Mean Percentage of Correct SA Probes (SE in parentheses)	154
Table 6. 6 Mean Probe Response Latency (SE in parentheses).....	155
Table 6. 7 Conflict Resolution Accuracy for the Correct and Failure Trials.....	155

List of Abbreviations

AC	Action Implementation
ADS-B	Automatic Dependent Surveillance-Broadcast
AIM	Impact on Mental Workload
ANOVA	Analysis of Variance
ARTCC	Air Route Traffic Control Center
ATMRI	Air Traffic Management Research Institute
AS	Information Analysis
ASRS	Aviation Safety Reporting System
ATC	Air Traffic Control
ATCO	Air Traffic Controller
ATC-SDK	Air Traffic Control-Sector Design Kit
ATM	Air Traffic Management
ATWIT	Air Traffic Workload Input Technique
CAAS	Civil Aviation Authority of Singapore
CANSO	Civil Air Navigation Services Organization
CDC	Clearance Delivery Controller
CD&R	Conflict Detection and Resolution
CDTI	Cockpit Display of Traffic Information
CR	Conflict Resolution
CRA	Conflict Resolution Aid
DD	Dynamic Density

DOA	Degree of Automation
ECG	Electrocardiograms
EEG	Electroencephalograms
EMG	Electromyograms
ERP	Event Related Potentials
EVSD	Electronic Vertical Situation Display
FAA	Federal Aviation Administration
FF	First Failure
FL	Flight Level
FMS	Flight Management System
FPS	Flight Progress Strips
HAI	Human-Automation Interaction
HITL	Human In The Loop
HRV	Heart Rate Variability
HSD	Horizontal Situation Display
IA	Information Acquisition
ICAO	The International Civil Aviation Organization
ILS	Instrument Landing System
IRB	Institutional Review Board
ISA	Instantaneous Self-Assessment
LAT	Look-Ahead Time
LOA	Level of Automation
LOS	Loss of Separation

LTM	Long Term Memory
M	Manual Condition
MAP	Missed Approach Point
MM	Mental Model
MP	Mental Picture
MSL	Mean Sea Level
NARSIM	NLR Air Traffic Control Research Simulator
NASA-TLX	The National Aeronautics and Space Administration-Task Load Index
N-CAID	NextGen Controller-Automation Interaction Display
NLR	Netherland Aerospace Laboratory
NPW	Number of Prior Warnings
OOTLUF	Out of the Loop Unfamiliarity
R	Reliable Condition
RA	Resolution Advisories
RAMS	Resolution Aircraft and Maneuver Selector
RSME	Rating Scale of Mental Effort
RT	Reaction Time
SA	Situation Awareness
SAF	Singapore Air Force
SAGAT	Situation Awareness Global Assessment Technique
SART	Situational Awareness Rating Technique
SATI	SHAPE Automation Trust Index
SHAPE	Solutions for Human-Automation Partnerships in European ATM

SID	Standard Instrument Departure
SPAM	Situation Present Assessment Method
SSD	Solution Space Diagram
SST	Situation Specific Reliability
STAR	Standard Arrival
STCA	Short Term Conflict Alert
SVS	Synthetic Vision System
TBE	Time Between Events
TBOCR	Trajectory-Based Conflict Resolutions
TCAS	Traffic Alert and Collision Avoidance System
TIS	Traffic Information Service
TLI	Task Load Index
TRACON	Terminal Radar Control
TSAFE	Tactical Separation Assured Flight Environment
UAH	Unreliable and High Time Pressure Condition
UAL	Unreliable and Low Time Pressure Condition
URET	User Request Evaluation Tool
USS	United States Ship
VASD	Vertical Separation Assistance Display
VSD	Vertical Situation Display

Summary

Air Traffic Control (ATC) safety is currently being threatened by the continuous increase in air traffic density (Airbus, 2013; Albrecht, Lee, & Pang, 2012; EUROCONTROL, 2004; ICAO, 2006, 2010; Sheridan, 2006). The continuous increase in air traffic could potentially trigger more Loss of Separations (LOS) as current ATC systems are approaching maximum capacity and the existing ATC practices may not be able to sustain the imminent traffic growth (CANSO, 2012). As a pertinent evidence, in 2004, 2009, 2013, and 2014 there were 144, 72, 112, and 163 cases respectively of near/mid-air collisions (U.S. Department of Transportation, 2014).

An automated Conflict Resolution Aid (CRA) has been proposed to be a plausible solution for providing an additional safety layer in the ATC. The automated CRA does not only have the function of detecting conflicts, but can also recommend controller maneuvers to eliminate the conflict. However, questions regarding suitable Human-Automation Interaction (HAI) procedures and safety critical issues in using the automated CRA in ATC remain to be answered (Prevot, Homola, Martin, Mercer, & Cabrall, 2012). The first unresolved question inquires about how the CRA reliability influences the application of the CRA, since automation reliability is arguably the most important issue in HAI. The second one questions how increased traffic density would influence the effects of CRA reliability on the use of the CRA. The third unresolved question asks how Vertical Situation Display (VSD) could support the application of the CRA in ATC facilities since lack of understanding regarding vertical situation has become one of the major causes of aviation accidents. This thesis therefore aims to address these questions.

This thesis is comprised of three experimental studies corresponding to the three research questions raised above. Study 1 investigated the relationship between the CRA reliability and human performance variables, including task performance, workload, trust, dependence, and situation awareness. The results of Study 1 showed that reliable automation could support operators' performances. However, the present study also verified the benefit of the imperfectly reliable CRA on ATCOs' performances. Study 2 examined the interaction between the CRA reliability, traffic density, and ATCO performance variables. Study 2 supported the findings in Study 1 by showing that the novel CRA, whether perfect or imperfect, could improve ATCOs' performances. In addition, the results indicated that the greater benefits of the CRA were applicable in high traffic density conditions, answering the future ATC challenge. Study 3 investigated the effects of VSD and its integration with the CRA. The results of Study 3 empirically supported the integration of VSD into ATC facilities as indicated by positive effects of the display on ATCOs' human performance variables. In addition, providing the VSD when ATCOs work with the CRA could help offset the automation imperfection.

The research theoretically describes various factors influencing ATCOs' performances with the CRA that covers automation as well as environmental aspects, including automation reliability, traffic density, and vertical situation display. Furthermore, this research provides empirical evidence regarding the relationships among the factors for the application of the CRA. Collectively, this research has high practical relevance. Results from this research provided practical implications for the development of future ATC workplaces, including conflict resolution automation and VSD for ATC facilities.

Chapter 1

Introduction

1.1 Research Background

1.1.1 Air traffic control

Increase in air traffic density is becoming the most vital issue in air traffic management. The International Civil Aviation Organization (ICAO) forecasts that worldwide air traffic density in 2025 will be twice as much as that in 2006 (ICAO, 2006). Similar trends were also reported by Sheridan (2006). In 2008, more than 10.8 million flight tracks were recorded in the sky above the Asia Pacific Region (ICAO, 2010). In the United States and Europe, the annual air traffic growth rates are 4.5% and 5%, respectively (Albrecht et al., 2012; EUROCONTROL, 2004). Increased air traffic will potentially trigger more air traffic conflicts since current Air Traffic Control (ATC) systems are approaching maximum capacity and the prevailing ATC practices may not be able to accommodate the imminent traffic growth (CANSO, 2012). Thus, this heavily compromises airspace efficiency and safety (Hoekstra, Van Gent, & Ruigrok, 2002).

Most fatal aviation accidents occurred while aircraft were in cruise (Boeing, 2015). As evidence, there is a U-shaped trend of near/mid-air collision cases (U.S. Department of Transportation, 2014). In 2004, there were 144 cases of near/mid-air collisions. The cases decreased to 72 in 2009. However, the trend started increasing year by year and reached up to 112 cases in 2013. Moreover, the cases further increased to 163 in 2014. These data indicate a real challenge for the future of ATC.

ATC is an essential component of Air Traffic Management (ATM). The nature of ATC is tactical in executing separation procedures for traffic collision detection and avoidance (Hoffman, Mukherjee, & Vossen, 2011). In today's operation, ATC is typically handled by Air Traffic Controllers (ATCOs), whose goals are to provide a safe, orderly, and expeditious flow of air traffic in their controlled sectors.

1.1.2 Automation in air traffic control

Given the increase in air traffic, ATCOs' workloads will inevitably increase since they need to resolve more potential conflicts (Murphy, Albert, Chen, & Anderson, 2012). Increased workload has been widely reported to be a major risk factor leading to aviation accidents (Billings & Reynard, 1984). To reduce the likelihood of aviation accidents, automation has been implemented in current systems. This was considered to be an effective solution for minimizing the workload imposed on ATCOs (Endsley & Kaber, 1999). In fact, the existing ATC facilities have been equipped with an automatic conflict alerting system that can inform ATCOs of possible air traffic conflicts prior to the actual conflicts.

Based on the facts of the strong automation benefits, automation has been implemented ubiquitously with the assumption that it can eliminate human error by lowering operator's workload. However, this assumption has been proven to be incorrect (Wickens, Hollands, Banburry, & Parasuraman, 2013). Although some forms of error may be reduced with automation, yet, it can introduce new errors in other forms, and may ironically increase human mental workload rather than decrease it (Pritchett, 2009; Sarter, 2008; Wiener & Curry, 1980). Human-Automation Interaction (HAI) research has

revealed that the nature of cognitive tasks is changed by automation in a way that was unexpected and unanticipated by system designers (Parasuraman & Riley, 1997). This, in turn, makes the human role becomes more essential as automation becomes more dominant. Unfortunately, most of system development efforts are devoted to the technical context with little or even no attention given to the human user's characteristics (Wickens, et al. 2013). Hence, the challenge for the automation development is to design for joint human-automation performance (Lee & Seppelt, 2009; Wickens, et al. 2013).

1.1.3 Automated conflict resolution

The automated Conflict Resolution Aid (CRA), which is a system that goes beyond detecting conflicts by recommending controller maneuvers to eliminate potential conflicts (Prevot, et al., 2012; Trapsilawati, Qu, Wickens, & Chen, 2015), is a component of the Traffic Alert and Collision Avoidance System (TCAS). Currently, the CRA embedded in the TCAS is used only in airborne aircraft. To date, however, ATCOs have not been served by an automated CRA system, although such systems do exist on the flight deck.

Current human factors research has mainly focused on the existing conflict alerting systems in ATC facilities, especially the effects of such systems' imperfections on ATCOs' and pilots' performance (Wickens, et al., 2009). However, even with the assistance of the conflict alerting system, ATCOs still need to further consider possible resolution maneuvers and to issue the commands while mentally simulating the conflict situation and coordinating other traffic aircraft (Kontogiannis & Malakis, 2013; Rantanen & Wickens, 2012). Given the fact that air traffic density will have doubled in the next two decades (Airbus, 2013), the higher probability of air traffic conflicts will become one of

the main challenges in ATC operations. The automated CRA is proposed to serve as an additional safety layer in the ATC system (Prevot et al., 2012) to maintain safety and the capacity of future air traffic (Cabrall et al., 2014).

However, questions on suitable HAI procedures and safety critical issues in using the automated CRA in ATC remain to be answered (Prevot et al., 2012). Research has highlighted the importance of ATCOs' trust in and dependence on automation during the design and development stages of ATC tools (Hilburn, et al., 2006; Nijhuis, 1999). In addition, a better understanding of ATCOs' workload and performance as well as situation awareness while using the automation aid is also required for specifying appropriate ATCOs-automation interaction procedures during the operational process (Francis T Durso & Manning, 2008; Rantanen, 2004).

Very few studies have been conducted to investigate how the CRA would affect ATCOs from the human factors perspective. Prevot et al. (2012) have recently reported that the CRA could improve ATCOs' performance. They also stated that ATCOs tended to work interactively with the CRA as long as they had final authority. However, in Prevot et al.'s (2012) study, they only examined a perfectly reliable CRA.

In order to make automation more effective in enhancing safety in the future increased air traffic environment, better knowledge of how humans interact with an imperfect CRA automation must be acquired. The automation may be imperfect due to automation or environmental factors such as sensor precision, software bugs, and hardware failure. To the best of the author's knowledge, nobody has studied how ATCOs deal with the CRA imperfection and time pressure. However, automation is never

perfectly reliable in this probabilistic world and such information can lead to disasters. An example of a disaster due to automation imperfection is the USS Vincennes accident in 1988. In this incident, the USS Vincennes wrongly classified an Iranian civilian airliner as an approaching “unknown-assumed enemy” aircraft, and thus led the Vincennes’ crew to mistakenly shoot at the aircraft. All 290 passengers onboard perished in this accident (Craig, Morales, & Oliver, 2004). This demonstrated the crucial need for a careful examination of automation imperfection and its effects on the human operator.

Furthermore, since the ATC domain is highly tactical in nature, the issue of time pressure for air traffic controllers is essential to be examined. In the current practice of the conflict resolution process, ATCOs have to deal with high time pressure conditions. The current conflict detection automation provides ATCOs with a projected conflict alert around 90 seconds look-ahead time, while the TCAS is activated 45 seconds before the conflict (Thomas & Wickens, 2005). The average time of maneuver instruction transmission is around 18 seconds (Cardosi & Boole, 1991), leaving ATCOs with only up to 27 seconds to come up with a resolution maneuver while monitoring other traffic aircraft. This time pressure is a critical threat to air traffic safety, yet no study has provided empirical delineation for the issue of time pressure in the conflict resolution process.

In addition, it is widely reported that ATCOs’ performance is affected by air traffic density (Galster, Duley, Masalonis & Parasuraman, 2001; Metzger & Parasuraman, 2001, 2005, 2006; Prevot, et al., 2012; Prevot, et al., 2010; Willems & Heiney, 2002). The possibility of an automation failure event occurring in high air traffic density is not zero and this could lead to a serious consequence for safety, however knowledge on the

interaction between CRA reliability and air traffic density has not been presented in the literature.

The lack of knowledge about vertical maneuvers was suggested to be a cause leading to insufficient understanding of automation (Vakil, Midkiff, & Hansman, 1996), even though vertical maneuvering advice is preferred by ATCOs (Rantanen & Wickens, 2012) and by pilots (particularly under high time pressure) (Thomas & Rantanen, 2006; Thomas & Wickens, 2006). In the ATC domain, Vertical Situation Display (VSD) does not exist in the current system. As such, the VSD may be beneficial in providing ATCOs with more knowledge about the vertical profile of aircraft, thus helping them to resolve conflicts related to vertical geometry and maneuvering, since currently this cannot be directly observed from a radar display. Based on this possibility, the interaction effects of CRA reliability and VSD were also examined in this study in which the VSD, by supporting situation awareness, may offset or mitigate the effects of imperfection. Thus, there is a crucial need to further investigate the concept of automated CRA from the perspective of ATCOs.

1.2 Research Objectives and Hypotheses

In an attempt to address the identified research gaps, this research intended to empirically study the effects of underlying factors of conflict resolution safety, including automation reliability, time pressure, air traffic density, and the effect of vertical situation display on air traffic controllers' trust, dependence, performance, workload and situation awareness. In particular, the research objectives of this thesis are threefold:

1. To investigate the effects of CRA reliability and time pressure on ATCOs' performance from the human factors perspective.

To answer the objective with regards to CRA reliability, the following hypotheses were proposed:

- H1.1. ATCOs' performance and workload would be improved with the support of the CRA as compared to manually performing the conflict avoidance tasks.
- H1.2. Perfectly reliable CRA would improve the performance and workload compared to the unreliable one.
- H1.3. Unreliable CRA would still improve the performance and reduce workload compared to the manual condition.
- H1.4. a. ATCOs' SA would be lower under automated conditions than the manual condition.
b. SA would be higher in the unreliable conditions than in the reliability condition.
- H1.5. Trust would degrade in the unreliable conditions compared to reliable condition and ATCOs would be less likely to depend on the CRA in the unreliable conditions.

In terms of time pressure, the following hypotheses were proposed:

- H1.6. Performance would decrease and workload would increase when ATCOs had shorter time available to respond in the unreliable condition
- H1.7. ATCOs' SA would be higher with increased time pressure.

To examine the proposed hypotheses, an experimental study was conducted. The human factors measurements (Langan-Fox, Sankey, & Canty, 2009) including conflict resolution performance, mental workload, trust, dependence and situation awareness (SA) were examined in a manual condition, and three automated conditions. The latter set consisted of one reliable condition, and two unreliable conditions and the latter two in turn were examined under low and high time pressure. In the automated conditions, participants were provided with the CRA. The participants manually performed the ATC tasks in the manual condition, similar to the current ATC practice.

2. To study the interactions of air traffic density and CRA reliability on ATCOs' performance from the human factors perspective.

To address the objective, the corresponding hypotheses were proposed as follows:

- H2.1. Imposing the CRA automation would increase both performance and reduce workload (when compared to fully manual conflict resolution), given that controllers use (depend upon) the automation.
- H2.2. The benefits in (H2.1) would be amplified under high traffic load, when cognitive resources were scarcer.
- H2.3. The benefits of the CRA automation would be manifest across traffic densities even when it was imperfectly reliable, given that its reliability was greater than the threshold of automation assistance between 0.70 & 0.75 (Wickens & Dixon, 2007).

- H2.4. CRA automation, whether perfect or imperfect, would decrease situation awareness (SA) of the traffic situation.
- H2.5. This loss of SA in (H2.4) would be manifest in compromised performance on the occasion of the FF of imperfect automation
- H2.6. The effects in (H2.4) and (H2.5) would be amplified under high workload (high traffic load).
- H2.7. After automation had failed, the trust in the CRA would decline but the dependence would not decline, due to the high traffic load.
- H2.8. High time pressure should decrease performance, increase situation awareness, and increase workload.

To examine these hypotheses, an experiment with two times and three times higher traffic load than the current level was carried out, where dependence on automation might be expected to be considerably higher, thereby amplifying both its benefits and its costs when it errs. Participants with ATC experience controlled traffic manually and also used a CRA tool that was either fully or partially reliable under different traffic density conditions. When the CRA was unreliable, either high or low time pressure was imposed. A variety of performance and cognitive variables both when the CRA aid worked perfectly and imperfectly were examined.

- 3. To study the effects of vertical situation display and its interactions with CRA reliability on ATCOs' performance from the human factors perspective.

The corresponding hypotheses were:

- H3.1. Replicating prior research, it is predicted that the CRA, even if imperfect, will assist performance and reduce workload relative to unaided manual performance.
- H3.2. The reduction in CRA reliability from 100% to 80% will degrade conflict resolution performance; but not to a level below manual resolution performance.
- H3.3.
 - a. The VSD will improve performance.
 - b. The VSD will increase situation awareness.
- H3.4. CRA reliability and VSD support will interact, such that the cost of imperfect CRA performance will be diminished when the VSD is present. Because of the latter's support for SA.

To examine these hypotheses, an experimental study was conducted to examine the benefit of an additional display (i.e. a VSD) that provides ATCOs feedback on the aircraft state. In conflict resolution, a VSD may provide the ATCOs with comprehensive vertical trend information, and help ATCOs examine the traffic surrounding the predicted conflict in order to ensure that the resolution will not lead to another conflict with other traffic aircraft. The integration of the VSD into ATC facilities and whether it would benefit ATCOs in resolving conflicts has never been explored, in isolation, let alone in conjunction with a conflict resolution aid, whether the aid perfect or imperfectly reliable.

Thus in the experiment, manual resolution was contrasted, with both a 100% reliable and an imperfectly reliable (80%) CRA. All three conditions were replicated with and without a VSD. Four dependent variables were of particular interest: overall

performance, first-failure performance, situation awareness (because declines in SA have been shown to degrade the human response to failure), and workload.

1.3 Research Scope

This research limits its scope to the evaluations of CRA and VSD from the human factors perspective. It did not address the technical development of CRA and VSD for the application in ATC workstation as simulated historical ATC scenarios were used during the studies. In order to reveal the effects of automation reliability, experimental studies were carried out. In the experimental studies, the effects of automation reliability on human factors' variables were investigated under varied task conditions defined by different levels of traffic density, time pressure and display conditions.

1.4 Dissertation Organization

Figure 1.1 shows the structure of the dissertation. This dissertation is divided into seven chapters. Chapter 1 introduces the research background followed by the research objectives and hypotheses, research scope, as well as the structure of the report. In Chapter 2, literature on automation and human-automation interaction related to the ATC domain is reviewed. Chapter 3 presents the research framework. Chapter 4 presents the study conducted to examine CRA reliability and time pressure, followed by Study 2 in Chapter 5 to investigate the interactions of both factors with air traffic density. Chapter 6 presents Study 3 examining the CRA with a VSD. Chapter 7 summarizes the key

findings, contributions, integrated information across the three studies, and limitations of this dissertation research, as well as proposes future research problems.

Figure 1. 1 Structure of Dissertation

Chapter 2

Literature Review

2.1 Air Traffic Control (ATC)

2.1.1 The Overview of ATC

ATC primarily covers the tactical nature of actions in air traffic management (Hoffman et al., 2011). Tactical control is a circumstance where ATCOs' performances depend upon a kind of planning (Corker, 1999). Air traffic controllers (ATCOs) are typically responsible for a certain area to ensure that all aircraft passing the area are doing so in a safe, order, and efficient manner. This responsibility includes monitoring aircraft in a particular sector to get information about aircraft call-sign, altitude, latitude, route, and destination (Sanchez, Krumbholz, & Silva, 2009) to ensure the aircraft are on the right path (Kallus, Van Damme, & Dittman, 1999), maintaining separation among aircraft and detecting air traffic conflict (Erzberger, Paielli, Isaacson, & Eshow, 1997), as well as resolving air traffic conflict (Canton, Refai, Johnson, & Battiste, 2005).

In today's operation, ATC is divided into three main responsibilities: tower control, terminal radar control (TRACON), and the air route traffic control centre (ARTCC) or so-called en route centre (Francis T Durso & Manning, 2008; Nolan, 1990). The path among those three facilities is illustrated in Figure 2.1.

Tower control is mainly responsible for controlling departing and arriving aircraft around the tower area and can typically be monitored visually, except during bad weather (Nolan, 1990). There are several ATCO duties in tower control, which are the local controller, ground controller, clearance delivery controller, and flight data controller. The

local controller's duties concern runway separation and sequencing, landing clearances and information, and runway and approach light system operation. The ground controller is responsible for issuing instructions for aircraft taxiing to runways, for ground support vehicles, and for relaying this information to airport management. Clearance delivery controllers (CDC) must obtain and relay clearance information to pilots, guide pilots in taking-off, and hand off aircraft to TRACON. The flight data controller's tasks are to obtain aircraft clearance from ARTCC and pass it to the CDC, to send weather information to pilots, to send departure messages to ARTCC, and to aid other tower controllers by relaying required information.

Figure 2. 1 Air Traffic Control Communication Path (Durso & Manning, 2008)

After taking off, aircraft is handed off to TRACON, which will control the aircraft for a subsequent 30-50 mile range. ATCOs working in the TRACON facility are classified into two different responsibilities depending upon their tasks. A departure controller is responsible for ensuring aircraft reach the appropriate altitude before being handed off to ARTCC and for handing off the departing aircraft to ARTCC. In contrast, an approach controller is responsible for ensuring aircraft descend to the appropriate altitude before being handed off to the tower control facility, in particular the local controller, and for handing off the arriving aircraft to the tower control facility at its destination. Both the departure and approach controllers are responsible for maintaining separation among aircraft within their assigned airspace.

From the TRACON facility, departing aircraft will be handed off to ARTCC, which controls the aircraft at the high speed and high altitude phases of flight. En route controllers, ATCOs in ARTCC, are responsible for separating aircrafts within their sectors, which appear on radar display (or so-called radar controller/ R-side), coordinating projected actions, and updating flight progress strips (FPS) to record aircraft movement (or so-called radar associate controller).

The aircraft are then handed off to any adjacent sector, from one ARTCC facility to another ARTCC facility, depending upon their destinations. While approaching their destination, the aircraft will subsequently be handed off to the TRACON facility and eventually to the tower control at each destination.

2.1.2 Air Traffic Conflict

During air traffic control operations, airborne aircraft have to be separated from each other. Although the definition of separation varies, the general separation standard being utilized in current operation is 5 nautical miles (nm) laterally and 1000 feet vertically (Francis T Durso & Manning, 2008; Isaacson & Robinson III, 2001; Nolan, 1990; Perry, 1997). This standard is applied by means of a “virtual cylinder” with a radius of 5 nm and height of 1000 feet within which an aircraft is flying. When two or more cylinders overlap at any point, this reflects an air traffic conflict as shown in Figure 2.2.

Figure 2. 2 Air Traffic Separation

The variation of separation standard may depend upon flight level and coverage range. Vertical separation standard is actually varied depending on the flight level of aircraft. For aircraft flying below 29,000 feet above mean sea level (MSL), the separation is similar to the current general separation standard, but for aircraft flying above 29,000 feet MSL, the vertical separation between aircraft is 2000 feet (Nolan, 1990). Lateral

separation standard is also varied, with 3nm for a less than 40nm radar coverage range and 5nm for a more than 40nm radar coverage range (Nolan, 1990).

ATCOs usually monitor aircraft movement within their sectors on a 2D radar display. They have been equipped with display supports to assess an aircraft's current position, to observe the aircraft's movement, and to predict the aircraft's trajectory. They will also have access to information detailing wind direction, individual aircrafts' performance capabilities, aircrafts' destinations and routes, and so forth. Once ATCOs observe adverse weather in a particular area within their sector, they may instruct the pilot to deviate from the flight plan trajectory and command the pilot to return to the original trajectory accordingly.

Since uncertainties are inevitable in a real and probabilistic world and ATCOs have to control many aircraft while maintaining the overall aircraft configurations, ATCOs may encounter air traffic conflict. In the current system, they could manually detect potential conflict in advance or an automation aid could detect the conflict first and alert the ATCOs. When an imminent conflict appears on the radar display, ATCOs must issue a command to the pilot in order to avoid the conflict.

2.2 Human-Automation Interaction (HAI) in ATC

2.2.1 The Overview of HAI

HAI is defined as a condition in which people (i) provide input, goals, and constraints to automation to be processed, (ii) control or supervision of the automation process, and (iii) receive automation outputs, which can be in many forms, including in the form of information (Sheridan & Parasuraman, 2005).

In ATC operations, ATCOs have to interact with highly sophisticated automation, which is designed to support the ATCOs in performing their work. There are three primary reasons for the application of automation in air traffic control systems: the requirements for enhanced safety and efficiency that include system flexibility, cost savings, and reduced personnel; the presence of the technology; and the needs to provide supports for ATCOs (Wickens, Mavor, Parasuraman, & McGee, 1998).

ATCOs work in different workspace settings depending upon their duties (Wise, Hopkin, & Garland, 2010). Some ATCOs work in the control tower and can see aircraft and other entities directly in clear and fair weather. Other ATCOs working in the terminal center and en-route center cannot directly view the aircraft they are controlling. ATCOs in these facilities work with data and computer assistance to identify aircrafts' position, flow, movement, environmental weather, etc.

Relationships between human and automation hold a very essential role as a means of allocating function to human or automation. The point of view of human-automation relationships has varied as technology advances. Initially, humans were forced to fit the automation, but eventually human and automation were seen as complementary entities, known as "human-centered" automation (Hopkins, 1991; Wickens, et al. 1998).

Human-centered automation covers the steps from automation purpose, development, evaluation, introduction, to operational application. First, the automation purpose or the decision of what to automate should focus on the compensation for human vulnerabilities and the exploitation of human strengths. Second, during the development of the automated tools, the users and trained human factors practitioners should be

actively involved. Third, for the evaluation of automated tools, human-in-the-loop simulation should be carried out with careful experimental design. Fourth, the application of the tools into the real workplace should be introduced progressively considering training, facility differences, and user needs. Fifth, the experience from the initial operational application should be carefully examined, and the mechanism to quickly respond to the examination results should be made available (Wickens, et al. 1998).

A key role for humans in automated systems is “supervisory control”, wherein humans supervise the automation and issue instructions regarding goals, constraints, and plans, and automation execute the instructions using their own systems with sensors and other sources (Sheridan & Parasuraman, 2005). However, there are various processes that can be automated (i.e. automation stage) which include information acquisition and integration; decision and action selection; and action implementation (Wickens, et al. 1998; Parasuraman, Sheridan, & Wickens, 2000). Furthermore, within these processes, automation can be considered in different forms that vary across a continuum of levels (i.e. automation level) (Wickens, et al. 1998). Collectively, the metric integrating these two dimensions (automation stage and level) represents the degree of automation (DOA) (Manzey, Reichenbach, & Onnasch, 2012; Onnasch, Wickens, Li, & Manzey; 2014). More reliable automation leads to greater trust in automation (particularly the latter) (Wickens & Dixon, 2007). More trust and a higher DOA both will often lead to more dependence or utilization (Lee & See, 2004). More trust in the automation is a cognitive state which may influence dependence on the automation, which is the behavioral manifestation of the trust. However dependence can also increase with higher workload even when trust is not there.

Higher dependence on the automation support will lead to lower workload (Loft, et al. 2007) and better routine performance (Onnasch, Wickens, Li, & Manzey, 2014). 2014), assuming that the automation is better than the unaided human. However the greater dependence or utilization causes a loss of SA because operators stop monitoring it (complacency) and stop generating actions themselves (loss of the generation effect), and because operators sometimes follow the advice of automation without thinking (automation bias: Mosier Skitka et al, 1998; Parasuraman & Manzey, 2010; Wickens, et al. 1998; Wickens, et al. 2013). This loss of SA causes real problems when automation errs (Li, Wickens, Sarter, & Sebok, 2014).

It is recommended that reliable, high-level automation applications should be provided for the stages of information acquisition, integration, and presentation and mid-level for decision selection in supporting all system functions as described in detail in the next section (Wickens, et al. 1998). Moreover, essentially important automation to be developed in the near future is the decision aids for *conflict resolution* and maintaining separation (Wickens, et al. 1998; Prevot, et al. 2012; Trapsilawati, et al. 2015). The direction of the development should ensure proper separation between aircraft during all phases of flight in order to improve flight path safety and efficiency (Wickens, et al. 1998). Yet, careful empirical human-performance evaluation on the concept of conflict resolution aid has not been well addressed.

2.2.2 Human Factors Concepts in HAI

2.2.2.1 Stage and Level of Automation

Function allocation between human and automation cannot be an “all-or-none” concept with a binary relationship (Kennedy, 2005; Parasuraman, Sheridan, & Wickens,

2000). Automation can be applied at different levels of autonomy, from fully manual to a fully automatic level depending upon the nature of the system. Table 2.1 presents levels of automation, whereby a higher level represents higher automation autonomy. For example, at level 4, automation will only suggest one alternative and will leave the decision to modify or agree to the alternative up to the human. At level 8, automation will inform the human only if the human asks; otherwise, the automation will not provide information on its actions.

Table 2. 1 Levels of Automation of Decision and Action Selection (Parasuraman, Sheridan, Wickens, 2000)

High	10. The computer decides everything, acts autonomously, with no human inputs
	9. Informs the human only if the computer decides to
	8. Informs the human only if asked
	7. Executes automatically, informs the human when necessary
	6. Allows the human a restricted time to veto before automatic execution
	5. Executes the suggestion if the human approves
	4. Suggests one alternative
	3. Narrows the selection of the alternatives down to a few
	2. The computer offers a complete set of decision/action alternatives
Low	1. The computer offers no assistance; human must make all decisions and actions

Extending the levels of automation, a model of the stages of automation was proposed (Wickens, et al. 1998; Parasuraman, et al. 2000). The model (see Figure 2.3) consists of four stages: (i) sensory processing, (ii) situation assessment, inferences, and diagnosis, (iii) decision making, and (iv) response selection. These four stages generally reflect “human information processing stages”. Sensory processing refers to information

input from sensory receptors, such as, in particular, selective attention. The second stage involves the use of the processed information in working memory to form an inference or situation assessment regarding the state of the world. The third stage involves cognitive efforts in decision making (Mosier & Fischer, 2010) and the last stage covers execution of the selected action(s) corresponding to output from the third stage.

Figure 2. 3 Stages of Automation

Endsley and Kaber (1999) and Proud, Hart, and Mrozinki (2003) developed other models representing function allocation between humans and automation (computer). Endsley and Kaber (1999) defined four generic functions that can be allocated to humans and/or computers: monitoring, generating, selecting, and implementing. These four functions were crossed with ten levels of automation (LOA) resulting in an LOA taxonomy as demonstrated in Table 2.2.

Table 2. 2 LOA Taxonomy for Human-Automation Performance in Dynamic, Multitask Scenarios (Endsley and Kaber, 1999)

Level of Automation	Functions			
	Monitoring	Generating	Selecting	Implementing
1. Manual Control	Human	Human	Human	Human
2. Action Support	Human/Auto	Human	Human	Human/Auto
3. Batch Processing	Human/Auto	Human	Human	Automation

4. Shared Control	Human/Auto	Human/Auto	Human	Human/Auto
5. Decision Support	Human/Auto	Human/Auto	Human	Automation
6. Blended Decision Making	Human/Auto	Human/Auto	Human/Auto	Automation
7. Rigid System	Human/Auto	Automation	Human	Automation
8. Automated Decision Making	Human/Auto	Human/Auto	Automation	Automation
9. Supervisory Control	Human/Auto	Automation	Automation	Automation
10. Full Automation	Automation	Automation	Automation	Automation

The model of levels of automation could be applied for the design of a future ATC system. In dealing with the increase in traffic density, additional automation to equip current ATC facilities is needed. The automation supports are required for flight scheduling and planning, traffic flow management, conflict detection, and conflict resolution. In ATC, conflict avoidance system includes conflict detection and conflict resolution. The conflict detection process in the current ATC system has been aided with a conflict alert that warns ATCOs of a pending imminent conflict. The conflict alert integrates airspace trajectory information, provides airspace inferences and its diagnosis of potentially intersecting paths, and thus falls into the Stage 2 of automation. Conflict resolution, however, involved the processes of developing and instructing a manoeuvre to be executed by pilots to avoid a detected conflict. The CRA, the automation that aids these processes, chooses or decides on a resolution maneuver based on the airspace assessment; hence it falls into the Stage 3 of automation that comprises decision and action selection.

A recommendation for the future ATC automation system was provided based on the level of automation (Wickens, et al. 1998; Parasuraman, Sheridan, & Wickens, 2000).

They suggested that a high level of automation could be applied for information acquisition and analysis to support SA. For decision and action execution, the high level of automation is only suitable for low-risk situations. For high-risk situations, the automation level should not exceed the capacity of providing preferred alternatives or suggestions to ATCOs.

Manzey, Reichenbach, and Onnasch (2012) provided empirical evidence to support the recommendation of automation levels and introduced the term of DOA which integrates the two dimensions (i.e. level and stage of automation). Higher levels and later stages of automation represent increase in DOA (Onnasch, et al. 2014). Manzey, Reichenbach, and Onnasch (2012) conducted an experiment on supervisory control tasks. They manipulated the DOA, into four conditions: the manual condition, information acquisition support (IA support), information analysis support (AS support), and action implementation support (AI support). Consistent with the recommendation, the highest DOA support showed a slight performance decrement. This was due to fewer verification steps taken by the AI support group for most complex failures. Given the highest DOA support, most participants tended to omit the required additional control actions. This can lead to adverse conditions, particularly in high risk situations.

Onnasch, et al. (2014) performed a meta-analysis using the DOA and observed patterned phenomena regarding routine performance, off-nominal performance (performance when dealing with an unexpected automation failure), and situation awareness. The meta-analysis' results showed that the higher the DOA, the better the routine performance. This finding is logical since automation works as how it is intended, to help human operators. However, as DOA increases, performance loss also increases

progressively when automation errs and human operators must do the task manually. Collectively, these two phenomena indicate a trade-off between better routine performance with higher DOA and worse performance when a failure occurs. This trade-off is associated with the progressive loss of SA that is observed with higher DOA and often known as “out-of-the-loop-unfamiliarity” (OOTLUF). The OOTLUF is a term related to Ergonomics field that occurs as a result from the combinations of complacency, reduced active engagement with the system, and progressive skill loss (Hancock, 1999; Sebok & Wickens, 2016).

2.2.2.2 Trust

Trust is a psychological construct with relevance to not only human-human interaction, but also in the human-automation interaction domain. Trust is defined as “*the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability*” (Lee & See, 2004, p. 54). This construct becomes relevant in the context of human-automation interactions given the extensive automation supports provided by the advancement of technology in everyday life. For example, a pilot will trust the cockpit display traffic information (CDTI) to show surrounding traffic information when adverse weather conditions occur or when other aircraft cannot be seen directly in the pilot’s line of sight.

Trust can be seen as an effect (i.e “trust as effect”) since there are some factors that influence it (Hoff & Bashir, 2015; Masalonis & Parasuraman, 1999). Automation performance or so-called reliability holds a substantial role in influencing humans’ degree of trust in something. This notion also fluctuates along the time depending upon human

experience. Lee and See (2004) proposed a trust-calibration model to illustrate the change in trust given automation capability and its perceived performance (see Figure 2.4). Well-calibrated trust is defined as a state where automation capability and trust level are equal. In contrast, less-calibrated trust refers to inequality between automation capability and trust level. Less-calibrated trust may initiate a decrease in trust level into either overtrust or distrust.

Figure 2. 4 Trust Calibration Function (Lee and See, 2004)

Trust calibration has three measures: perceptual accuracy, perceptual sensitivity, and trust sensitivity (Stephanie M Merritt, Lee, Unnerstall, & Huber, 2014). Linked to the measures, Hoff and Bashir (2015) proposed three different layers of trust: dispositional trust, situational trust, and learned trust.

The first layer, dispositional trust, is based on human-related factors and reflects individual's overall tendency to trust automation, independent of context or a specific

system. This first layer is arguably related to the perceptual accuracy that indicates an initial user's perception of automation reliability.

The second layer, situational trust, places an emphasis on the development of trust that varies depending upon the situation (environment-related factors). This second layer also reflects the perceptual sensitivity where a user's perception of automation reliability varies in correspondence to actual reliability changes.

The third layer, learned trust, considers automation-related factors. It represents an operator's evaluations of a system based on the previous or the current interaction with automation. This last layer is associated with trust sensitivity that indicates the trust changes as the actual reliability changes. Merrit, et al. (2014) found that perceptual accuracy is the only calibration measure that was highly associated with task performance.

However, based on a meta-analysis, Schaefer, et al. (2016) only found medium effects of both human- and automation-related factors on trust. Further analyses on the factors revealed that, however, human-related emotive factors and automation-behaviour play more important roles than environment-relative factors in the development of trust.

Trust is not only understood as the most important construct in human-automation interaction (Wickens, et al. 2013), but it is also a key point for future ATC, particularly in the ATCO's role (Langan-Fox, Sankey, & Canty, 2009). Trust has increasing importance for future ATCO roles since automation supports will be ubiquitous in future ATM systems. However, very few studies have been carried out to investigate this substantial construct, especially in the ATC domain (Kelly, Boardman, Goillau, & Jeannot, 2003; Langan-Fox et al., 2009).

Only few existing trust measures are available. Majority of trust measures were a subjective approach. Since trust is a notion at a cognitive/affective state, subjective rating is arguably the most appropriate technique to measure it (Wickens & Xu, 2002). SHAPE (Solutions for Human-Automation Partnerships in European ATM) Automation Trust Index (SATI) is an example of trust measures which is used in ATC domain. The subjective measure was developed for assessing automation tools in an ATC system and the questions were established based on ATCOs' inputs (Goillau, Kelly, Boardman, & Jeannot, 2003). Madsen and Gregor (2000) developed a questionnaire to measure trust in human-computer interaction using five underlying principles which are reliability, understandability, technical competence, faith and personal attachment. Lee and Moray (1992) developed their own subjective trust rating by providing a persuasive explanation of trust definition and this technique was also adapted by other studies. Another validated questionnaire to assess trust was created by Jian, Bisantz, and Drury (2000) and this questionnaire has successfully placed trust and distrust as two opposite ends of a single continuum (Atoyan, Duquet, & Robert, 2006). SATI was developed to address more on ATC computer assistance tools.

SATI, Human Computer Trust Rating Scale (Madsen and Gregor, 2002), and Trust Subjective Rating (Lee and Moray, 1992) are easy to be applied, however the trust conceptualization as a binary construct is not clear. The questionnaire developed by Madsen and Gregor (2000) also showed an inclination of individual bias. Among these available questionnaires, the questionnaire developed by Jian, Bisantz, and Drury (2002) is seen as the most appropriate questionnaire to be used in the experiment since it has

successfully verified the trust and distrust constructs as two-opposed ends of a single continuum.

Kaniasaru, Steinfeld, Desai & Yanco (2012) introduced an objective measure for trust. They proposed two potential measures to be indicators for trust loss. The first one is Time-Between-Event (TBE) which provides the time between two consecutive alerts indicating the time between the most recent warning and a participant changing of control mode. The second measure is Number-of-Prior-Warnings (NPW) which specifies the frequency of warning experienced by participant prior to change of control mode. This measure may have a higher validity in measuring trust. However, this measure can only be applied at a certain case (i.e. alerting system with mode changing system as a backup solution). In addition, this method cannot measure initial trust.

Taken together, trust subjective rating method developed by Jian, Bisantz, and Drury (2000) is seen as the most appropriate approach to measure trust. This is because trust as a cognitive construct can be more appropriately assessed using a subjective rating. Furthermore, Jian, Bisantz, and Drury's (2000) method have gained the highest validity to measure trust and distrust as two opposed ends of a single continuum.

2.2.2.3 Dependence

It is important to differentiate between trust in automation and dependence on automation. While trust is a notion at the cognitive/affective state of the user that can be assessed by subjective ratings, dependence is the behavioural manifestation of trust that

can be assessed through the interaction between the user and the automation (Wickens, et al. 2013).

Dependence reflects the extent to which the human operator uses the automation, follows its advice, and verifies the automation's recommendations based on the raw data (Bahner, et al. 2008; Wickens, et al. 2013).

There are two different forms of automation dependence for stage 1 and stage 2 diagnostic systems: reliance and compliance. Reliance reflects the extent to which an operator reduces monitoring, active responding and engagement with the automation system when the automation is "silent" (the alarm does not activate) (Meyer, 2001; Sebok & Wickens, 2016). The automation operator exhibits high reliance behaviour when he or she stops monitoring the raw data while the automation alarm is not activated and it informs the user that everything is all right (Wickens, et al. 2013; Dixon & Wickens, 2006; Meyer, 2001; 2004).

Compliance refers to a situation when a human reacts promptly due to alerts or alarms from a monitored system when the alarm activates (Dixon & Wickens, 2006). An operator shows high compliance behaviour if he or she always follows or immediately follow all automation suggestions or actions (Lee, 2008; Parasuraman & Riley, 1997) without checking raw data from the environment (Bahner, Hüper, & Manzey, 2008) due to over-trust in the automation's function (Parasuraman & Manzey, 2010), as closely related to the automation bias (Mosier & Skitka, 1996). Operators show a lack of compliance when they fail to respond to an alarm, thinking that it must be a false alarm.

In contrast to dependence, however, human operators often reflects sceptical behaviour towards automation, which triggers a human to take manual action due to lack

of trust in automation (Lee, 2008; Parasuraman & Riley, 1997), and to ignore the advice of automation, whether that automation provides guidance as to the state of the world, or recommends an action. In addition, they may simply reject the automation because of their pre-existing knowledge or emotions regarding past experiences with automation (Hoff & Bashir, 2015).

Nevertheless, trust and dependence are not always in a linear relationship. In some cases, such as overloading or executing very complex tasks, humans may rely on the automation even if they do not trust the automation (Atoyan et al., 2006). Furthermore, imperfect automation may be depended upon, even if it may be mistrusted, because of the high tasks' workload (Wickens & Dixon, 2007) or lack of awareness of the automation's limitation (Hoc, 2001).

The relationship between trust and dependence can be determined by many factors, such as automation complexity, situation novelty, operator's awareness of manual and automated performances, and operator's freedom to make decisions (Hoff & Bashir, 2015). Hence, trust and dependence should be measured separately as two different entities in order to gain clearer picture of the effects of an automation manipulation on the two constructs.

2.2.2.4 Situation Awareness

SA is defined as the condition wherein humans are aware of what is going on. The formal definition of situation awareness is "the perception of elements in an environment, within a volume of space and time, and comprehension of their meaning and projection of

their status in the near future” (Endsley, 1995, p. 36). Based on Endsley’s definition, SA was hence categorized as level 1, level 2 and level 3.

The importance of SA to HAI is clear in the sense that losing SA will compromise safety due to OOTLUF (Wickens, et al. 2013). This unfamiliarity could be a serious matter when dealing with unexpected dangers where ATCOs are unable to take appropriate actions. This is particularly true in the monitoring task as ATCOs’ SA was found to be lower and even lower when given a higher task load (Willems & Truitt, 1999). Research also suggests that a high DOA automation system may not be appropriate since it could cause ATCOs to become out of loop and thus threaten SA. Therefore, a future automation system must be designed to address these issues and to maintain ATCOs’ SA at an appropriate level that prevents OOTLUF (Langan-Fox, Sankey, and Canty, 2009).

In regards to measuring SA, Langan-Fox, Sankey, and Canty (2009) provided a comprehensive review about the various methods used in assessing SA in ATC. Both subjective and objective approaches are available to assess SA. Situation Awareness Rating Technique (SART) is an example of the subjective approach because it asks the operator if they think they have SA. This method removes secondary task processing and non-intrusive. However, it reflects limited dimensions of SA, limited validity/sensitivity of the construct’s scaling, and it cannot provide SA data throughout trial. The Situation Awareness Global Assessment Technique (SAGAT) is an example of the objective approach because it asks the operator to answer an objective question about the situation, but this approach relies heavily on ATCOs’ workloads and is highly intrusive. Another method to measure SA is Situation Present Assessment Method (SPAM). SPAM consists

of ready response and probe response to allow more precise measurement of SA and differentiation of SA from ATCOs' workloads. This method is appropriate for ATC context since it uses prompt-type probes where ATCOs process information in this manner. It also has high validation and importantly, it is less intrusive than other objective methods. In addition, this method removes problems associated with post-trial data, is quick and easy to use with minimal training. The cons of this method are that it requires long preparation time and probes developed during task imposes burden on the researcher.

For physiological measure for assessing SA, eye tracker is used. This method is non-intrusive to task performance and can be used to determine environmental elements of interest. But the cons are that it does not measure quality of planning, it only provides indirect assessment of SA, and ultimately, it restricts operator's movements. If the ATCO moves (and s/he typically does, to see the flight plan data, to communicate with pilot and to coordinate with planning controller), data analysis of the SA will not be precisely reflect SA or even cannot be obtained. Among all of the methods that were described, SPAM is the method that has the most advantages and fewest disadvantages in an ATC context. SPAM does not burden ATCOs' memory, supports use in the ATC context, and is less intrusive (Bacon & Strybel, 2013). SPAM will be employed in the current research.

2.2.2.5 Performance

Human performance reflects various tasks from simple to complex ones that are accomplished by a human operator. Human performance is measurable and the measures can be classified into two main types. The first type is accuracy, which reflects "the degree of correctness" (Gawron, 2008). This type assumes that there are correct or

successful trials. The second type of human performance measure is time. This type assumes that the duration of task performances can be measured since the beginnings and endings of the tasks are defined, and that shorter time is better performance.

Accuracy in general is “a measure of the quality of behavior” (Gawron, 2008). Number of correct trials (e.g. conflict detection and conflict resolution trials) (Metzger & Parasuraman, 2001; 2005; 2006; Rovira, et al. 2010) and percentage of correct trials (Rovira, et al. 2007; Trapsilawati, et al. 2015) are examples of accuracy measures. Percentage of correct trials has been typically used to evaluate the effects of visual display, automated decision aids, training, and tasks. Accuracy can also be measured using error, for example through error rate (Hunt, 2013), number of errors (Goritzlehner et al., 2014), and false alarm and miss rate (Imbert et al., 2014; Sanchez, Rogers, Fisk, & Rovira, 2014). Errors are classified into two categories: omission and commission errors (Wickens, et al. 2013). Omission errors occur when a human operator completely ignores the task whereas commission errors occur when a human operator does the task incorrectly.

Some measures of time include reaction time (Jha, et al. 2011; Dao, et al. 2011), searching time (John & Manes, 2002), and completion time (Li, Wickens, Sarter, & Sebok, 2014). Reaction time is defined as the time taken from the onset of a stimulus to the action of a human behavior toward the stimulus. Searching time is the time required for a human operator to complete the searching task. Completion time is the time required to complete a task.

2.2.2.6 *Performance & Reliability in imperfect automation*

Humans can perform tasks manually or with the help of automation (i.e. monitor an automated aid with different degrees of intervention, please refer to Section 2.2.2). Onnasch, Wickens, Li, and Manzey (2014) differentiated performance into routine task performance and off-nominal performance. The routine task performance is a cooperative performance between the human operator and the automated system when the system functions properly. The off-nominal performance is performance when there was off-nominal situation due to automation or human response to unexpected automation failure (Sebok & Wickens, 2016), such as complete machine breakdown (Onnasch, et al. 2014), or due to environmental factors, such as thunderstorms or weather issues in an ATC context; situations that automation may not be designed to handle (Rovira, et al. 2010; Prevot, et al. 2012).

There was a tradeoff between routine task performance and return-to-manual performance based on the meta-analysis conducted by Onnasch, et al. 2014. They found that the higher degree of automation could support better routine performance but would worsen manual performance when a failure occurred. This phenomenon is induced as a result of OOTLUF: the combination of complacency, lower engagement with the automation system, and the loss of generation effect due to the progressive loss of skill (Sebok & Wickens, 2016).

The trade-off model showed the divergence of the routine task and return-to-manual performance curves as DOA increases (Onnasch, et al. 2014; Sebok & Wickens, 2016). This phenomenon was empirically supported by Li, et al. (2014), who found that automation can improve routine performance but leads to costs in the failure condition.

Compared to manually performing the tasks, operators working with automation performed better in the highly reliable condition. However, for the low reliability condition, performance benefit was associated with increased attentional efforts and a declining performance in concurrent tasks (Onnasch, Ruff, & Manzey, 2014).

The relation between the reliability of imperfect automation, routine, and failure performance is as follows: Overall performance of automation that is, for example, 80% reliable will be the weighted average of the performance of the frequent (80%) times that it is successful, and the rare (20%) of the times it fails. That is $\text{perf}(\text{overall}) = 0.8 \times \text{perf}(\text{normal}) + 0.2 \times \text{perf}(\text{fail})$. $\text{Perf}(\text{normal})$ will be better and $\text{perf}(\text{failure})$ will be increasingly worse as DOA increases. But generally, the benefits of normal automation (being correct most of the time) outweigh the costs of failure automation (being wrong, but rarely). This results in an overall benefit of imperfect automation as long as the 80/20 ratio (or higher) is preserved (Wickens & Dixon, 2007). However the one caveat to this general description is that the very first time that automation fails, (“first failure”) this failure performance may be a lot worse, than on the subsequent failures by which time where people will have learned to retain certain skepticism in the automation or to calibrate their trust appropriately to the 80% level.

When dealing with a system’s error, different operator-system actions can lead to different consequences. The consequences may also vary depending upon the nature of the system. Hence, human performance when given imperfect automation is a crucial area to be examined, particularly for the *state-of-the-art* concept of conflict resolution automation.

2.2.2.7 Workload

In an ATC system, ATCOs' workload has been observed to greatly influence handling capacity (Hopkin, 1991). Thus, this construct becomes one of the most important concerns related to ATC (Rantanen & Nunes, 2005). Although the definition of workload covers both physical and mental workload, the majority of ATC task demands are associated with mental workload. ATCOs generally have two main goals, which are to maintain separation based on ICAO standards and to control the aircraft in a safe and orderly manner until they reach their destination. These goals require ATCOs to perform multiple and simultaneous tasks, including monitoring, anticipating, and intervening. These task demands that are associated with mediating factors are found to affect ATCOs' mental workload (Loft, Sanderson, Neal, & Mooij, 2007).

Task demands in ATC are influenced by air traffic factors that are defined in terms of traffic number and complexity. Traffic number reflects density through aircraft counts in a particular sector and their vertical and horizontal distances between one another. Complexity factors are classified into two categories: aircraft transition factors and possible conflicts in a sector (Loft, et al., 2007; Murphy, et al., 2012). Correspondingly, the mediating factors are related to ATCOs factors such as skill, experience, and strategy, as well as to the automation factors such as quality of equipment (Hilburn & Jorna, 2001; Mogford, et al. 1995; Loft, et al. 2007). The generic form of an open-loop mental workload model is illustrated in Figure 2.5.

Figure 2. 5 Open-Loop Model of ATCOs' Mental Workload (Loft, et al. 2007)

Several workload measurement methods have been introduced: behavioural measures, secondary task, subjective measures (Wickens, et al. 2013), predictive model (Averty, et al. 2004), and physiological measures (Wickens, et al, 2013). First, workload can be measured effectively by examining the human behaviour which is defined as the form of performing something along the time (i.e. time line analysis). Timeline analysis is an effective behavioural measure of workload that may change depending upon task demands even when performance remains constant. However, it can only effectively measure workload in manual control tasks.

Second, workload can be assessed through secondary tasks performance and it reflects objective workload. The rationale of this measure is that with more residual capacity from the primary tasks, better secondary tasks can be performed, which also infers lower resource demands for the primary task (Wickens, et al, 2013; Boag, et al. 2006). However, the main challenges with this measure is the intrusiveness where researchers may not be able to control the amount of attention devoted to the secondary tasks and for secondary task performance, some tasks may be not related to the demands of the primary task (Rehmann, 1995). The examples of secondary tasks performance measures for assessing workload are ready response latency and percentage of timeouts. Researchers can control the attention for the secondary tasks with these methods because

of the approach for non-responded question that allows participants to not tunnel their attention. In addition, the tasks are also related to the primary task.

Third, subjective measures of workload (perceived workload) are techniques that have been widely used. The subjective approach is typically presented in a questionnaire format asking humans to subjectively rate their perceived workload level. The majority of workload reports in the ATC domain relied on subjective assessment (Hilburn & Jorna, 2001 in Langan-Fox, Sankey, & Canty, 2009). NASA Task Load Index (NASA-TLX), Air Traffic Workload Input Technique (ATWIT), Instantaneous Self-Assessment (ISA), Impact on Mental Workload (AIM), and Rating Scale of Mental Effort (RSME) are examples of subjective approaches (Langan-Fox et al., 2009).

NASA-TLX is a reliable measure (Endsley and Kaber, 1999) and is widely validated; hence it is used extensively in aviation research (Young, Brookhuis, Wickens, & Hancock, 2014). The disadvantage of the NASA-TLX is that it does not consider specific activities that can contribute to mental demand. The advantage of ATWIT is that participants do not need to separately indicate the factors of mental workload. However, this measurement method can increase a controller's perceived mental workload since the controller should respond in a real-time manner. ISA is reported to be a valid measurement; however, it is not indicative. AIM is easy to implement in a real time simulation, however its validity and reliability are still questioned (Rehmann, 1995). RSME has a high face validity but its inter-subject variability is very high (Rehmann, 1995). Out of all of the existing workload measurement methods, however, it can be inferred that NASA-TLX is the most commonly utilized and the most extensively

validated technique in comparison to other techniques (Salmon, Stanton, Walker, & Green, 2006).

Fourth, workload can also be measured by the predictive modelling of task load variables that cause workload. For example, a traffic load index (TLI) that is defined as traffic load within time boundaries and is computed by summing up all of the events occurring during the several time units measured between time boundaries (Athènes, Averty, Puechmorel, Delahaye, & Collet, 2002; Averty, Collet, Dittmar, Athenes, & Vernet-Maury, 2004). The TLI method can be used for assessing workload in real-time environments, but it is not easy to apply since it requires ATCOs' expertise (Athènes et al., 2002).

Lastly, workload can be assessed using physiological measures, such as Heart Rate Variability (HRV) (Metzger & Parasuraman, 2006), pupil diameter, visual scanning, entropy, Event-Related Potentials (ERPs), Electrocardiograms (ECG), Electroencephalograms (EEG), and Electromyograms (EMG) (Langan-Fox, et al. 2009; Wickens, et al. 2013). These measures are able to detect workload fluctuation during a short time period. However, these measures are complex, costly, and less sensitive than subjective measures. As a result, these measures are less frequently used in the dynamic ATC environment (Langan-Fox et al., 2009).

Given the higher advantages of the subjective approach in measuring mental workload, this approach was used in the current study. Due to the practicality, reliability, and validity of the NASA-TLX, this method was preferred in measuring mental workload for the experiments. In addition, ready response latency and percentage of timeouts were used to objectively measure the workload.

2.2.3 Drivers of HAI in ATC

There are various factors influencing the interaction between ATCOs and automation. Of all the drivers of HAI in ATC, automation reliability is particularly the most critical one (Wickens, et al. 2013), given the tactical nature of ATC (Vossen, Hoffman, & Mukherjee, 2012). In addition, the development of ATC automation is evolving due to the higher traffic demands and conflict resolution (CR) automation is one of the most recent promising ATC automation concepts that is in undergoing research.

The following section reviews studies relevant to the factors influencing the applications of ATC-related automation. The studies can arguably provide pertinent fundamentals for the new concept of CR automation, regarding how does the reliability of CR automation affects ATCOs' performance and SA with increasing workload due to traffic density and time pressure. In addition, when CR automation fails (e.g. due to software or hardware problems), the issue of how are the costs of imperfections mitigated by automation displays is inevitably crucial.

2.2.3.1 Reliability

Human trust in automation is essentially affected by the automation reliability (Lee & See, 2004). There is a relationship between the true capabilities of automation and the level of trust. However, since perfectly reliable system is almost impossible in this probabilistic world, human trust will apparently be influenced by perceived system reliability, thus inducing different decisions to depend or not depend upon an automated system.

Reliability is typically defined as the ratio of an automation being correct in its operation to all operations performed. The automation is perceived as unreliable when, for

example, the automation recommends a wrong choice or executes a wrong action. In reference to the definition of reliability, a system is perceived to be 70% reliable if it can correctly perform 7 tasks out of 10 designated tasks (Wickens and Dixon, 2007).

Only two studies have been conducted to investigate the issue of automation reliability of conflict detection in the ATC domain. The first study (Metzger & Parasuraman, 2005) compared different conditions of conflict detection automation including reliable, automation failure, and manual conditions. The results showed that more conflicts were detected and the conflicts were detected earlier in reliable condition than in manual condition. However, when automation failed in one of the trials, ATCOs were more likely to detect the conflict in manual condition than in unreliable condition.

Rovira and Parasuraman (2010) examined the effect of imperfect automation on visual attention and conflict detection performance. They manipulated automation support variables into four levels, including manual, reliable automation, miss automation, and false alarm automation. They found out that automation increase accuracy and speed in the conflict detection process as long as it was reliable when compared to the manual condition (no automation support). They also discovered that imperfect automation could lead to a poorer conflict detection performance. Although not statistically significant, a tendency to minimize visual attention through a fewer number of fixations under the false-alarm condition was observed. This finding is supported by Manzey, Gerard and Wiczorek (2014) who found that low alarm reliability led to non-responses to a significant number of alarms.

However, analysis of real ATC data showed no evidence of negligence behaviour (Wickens, et al., 2009). Wickens, et al. performed an analysis of 495 conflict alerting

cases wherein forty-five percent of the data consisted of false alerts. Although ATCOs exhibited a propensity towards non-responses due to false alarms, no evidence was observed that ATCOs did not respond or delayed their responses to true alerts. This being said that negligence in the context of en-route ATC conflict alerts is not essentially affected by a high false alarm rate.

A number of studies, particularly within the aviation domain, have been carried out by manipulating the level of a system's reliability. In their aviation automation research, Bailey and Scerbo (2007) set a high level of system reliability where a 2% failure rate (i.e. 2 out of 100 deviations failed to be detected) was set to occur under high reliability conditions, and a 13% failure rate was set for low reliability conditions. The proportion of correct detections increased under the low reliability condition. This suggests that as the reliability level increases (perfect or nearly perfect), the tendency towards the "complacency" effect was higher. The operator response time also supported this finding where slower response time was observed under the high reliability condition than under the low reliability condition.

Parasuraman, et al. (1993) manipulated reliability levels into high and low levels with 87.5% and 56.25% system reliability, respectively. The former indicated that automation could detect 14 out of 16 malfunctions, while the latter reflected that the automation could detect only 9 out of 16 malfunctions. They set these two reliability levels into a variable-reliability group and a constant-reliability group. The results showed that better performance was observed in the variable-reliability group. Automation-induced complacency was apparent in the constant-reliability group.

The effects of automation reliability on trust and dependence showed some directional patterns, although the findings were not perfectly consistent. The majority of prior studies indicated that higher automation reliability would lead to higher trust (Metzger & Parasuraman, 2005; Dzindolet, et al., 2003; Lee & Moray, 1992; Merritt & Ilgen, 2008; Parasuraman & Miller, 2004; Wang, et al., 2009; Weinstock, Oron-Gilad, & Parmet, 2012). However, no significant effect of aid reliability was found to affect trust in command and control tasks (Rovira, McGarry, & Parasuraman, 2007).

In terms of dependence, Dzindolet, et al. (2003) found that participants using military automation showed an inclination towards relying on themselves over the automation when the automation made errors; however, no significant effect of reliability was found to affect reliance on the combat identification system (Wang, et al., 2009) and on the automation alerting system (Rice, 2009). Further, Merritt and Ilgen (2008) found that the tendency to trust the automation increased the prediction of the decision to use automation. Higher trust often led to higher dependence on the automation. Trust would degrade in the unreliable conditions in comparison to the reliable condition and therefore, participants would be less likely to utilize the automation in the unreliable conditions.

Wickens and Dixon (2007) provide a synthesis of the literature on imperfect diagnostic automation. Through a meta-analysis of 20 studies in various domains, they suggested that 72% was the “crossover” point for the reliability level of a system. Given an 8% margin of error in measurement, to be on the conservative side, they suggest that reliability should be $> 80\%$ ($=72\% + 8\%$). This indicates that system reliability below that crossover point will lead to a worse impact on human performance than no automation at all.

Collectively, the effect of automation reliability on ATCOs' performance has been consistently reported that reliable automation would lead to better routine performance. However, it could lead to adverse effects when it fails, due to loss of SA. Furthermore, there was a tendency that the trust in and dependence on automation will increase as the automation reliability increases.

2.2.3.2 Traffic Density

Traffic density is a driver of ATCOs' workload as discussed above in section 2.2.3.5. It has also been shown to affect ATCOs' performance and dependence on ATC-related automation (Galster, et al., 2001; Metzger & Parasuraman, 2001, 2005, 2006; Prevot, Homola, & Mercer, 2008; Prevot, et al., 2010; Willems & Heiney, 2002; Westin, Borst, & Hilburn, 2013).

Galster, et al. (2001) conducted a study using an ATC simulation task in a free-flight (FF) environment. Two conflicts and four self-separations were set in the scenario over two levels of traffic density, a moderate level and a high level, involving 11 aircrafts and 17 aircrafts, respectively. As expected, they found that high traffic induced a higher mental workload and objective workload in comparison to moderate traffic. Generally, ATCOs' performances, as indicated by the number of detected conflicts and self-separation and advanced notification times, were poorer under high traffic density than in a moderate traffic density environment.

Metzger and Parasuraman (2001, 2005, 2006) conducted three different studies involving similar traffic density manipulation. They manipulated traffic density into two

levels, moderate and high, with 10 aircrafts and 16 aircrafts, respectively. The first study (Metzger and Parasuraman, 2001) investigated traffic density and ATCOs' control mode (passive and active monitoring). Similar to Galster's (2001) finding, the result suggested that high traffic density led to a higher mental workload and poorer performance, as indicated by lower advanced notification times, particularly under the passive monitoring mode, which shows the cost of passive monitoring. The second study (Metzger and Parasuraman, 2005) examined traffic density and the effect of conflict detection aid (presence and absence of conflict detection aid). Similarly, the result showed that general performance was better with moderate traffic density than with high traffic density. Mental workload was found to be higher under high traffic density than under moderate traffic density through the use of heart rate variability measurements and subjective ratings. The third study (Metzger and Parasuraman, 2006) dealt with traffic density and cue types (simple visual cuing, enhanced visual cuing, and multimodal cuing with tone) in the ATC conflict detection domain within the free-flight environment. Interestingly, the result exhibited a slightly different trend than the prior two studies. No significant effect of traffic density on the advanced notification times for conflicts was observed. However, self-separations were detected earlier under moderate conditions than under high traffic density conditions. For mental workload, the result was similar to the two other studies in indicating higher workload under high traffic density than under low traffic density. No significant effect of traffic density on controllers' trust ratings was found.

Nunes (2003) conducted a study using predictive display and manipulated the display across two airspace loads: low load and high load. The results revealed that

response times were significantly longer with a high load than with a low load, particularly for unaided groups of controllers.

Willems and Heiney (2002) examined task load effect on the use of a decision support tool in an en-route environment. The task load effect was represented by both complexity and aircraft counts, which are an average of 25 aircrafts and 35 aircrafts for low and high loads, respectively. The findings were that faster ATCOs response time was observed when the task load was low. ATCOs rated their mental, physical, temporal workload, effort and frustration levels as being higher under a high task load than under a low task load scenario. ATCOs' trust in the decision support tool was found to be higher with a low task load. This might be because of the longer time available for ATCOs to try and work with the automation.

A study by Prevot, et al. (2010) was conducted to examine the effect of traffic density by manipulating the aircraft count to be 42 aircraft (level A) and 55 aircraft (level B). The results indicated that a higher traffic level induced significantly greater workload than a lower traffic level. ATCOs' performances were also adversely affected, as indicated by a higher average number of operational errors under level B than under level A.

Prevot, et al. (2008) carried out a study to investigate the effect of different levels of conflict resolution authority (manual, interactive, and fully automated) across two levels of traffic density with a throughput rate of 30 aircrafts and 60 aircrafts, respectively. The results showed that higher traffic was unmanageable using trial planning function alone, but was somewhat manageable using the interactive mode. A significantly higher number of conflicts and higher separation violations occurred under high traffic

than under moderate traffic. Although the result also indicated that controllers generally accepted the automation, there was a slight declining trend as traffic density increased. Workload was found to be higher under high density on the interactive mode, but it was still less than that of the manual mode.

Westin, et al. (2013) examined the effects traffic load and resolution strategy on ATCOs' dependence and workload. The results showed that participants accepted more resolution advisories in the high traffic load and the workload was significantly increased with more traffic.

In conclusion, the research on traffic density has so far been able to exhibit consistent findings in relation with workload and performance. Higher traffic density led to higher workload and thus led to poorer performance as indicated by lower conflict detection and resolution rates as well as longer response times. However, the effects of traffic density on trust and dependence are not consistent. For trust, Metzger and Parasuraman (2006) found no effect of traffic density on trust whereas Willems and Heiney (2002) found higher trust with lower traffic density. The results on dependence showed opposite directions - In the study of Prevot, et al. (2008), the trend to depend upon automation decreased as traffic density increased, while in the study of Westin, et al. (2013), the dependence increased as traffic density increased.

Furthermore, the effects of traffic density in conflict detection and resolution were only examined in the relation with presence and absence of automation (Metzger & Parasuraman, 2005; Prevot, et al. 2010), the control mode (Metzger & Parasuraman, 2001), the automation cueing (Metzger & Parasuraman, 2006), the automation intervention level (Prevot, et al. 2008), and the resolution strategy (Westin, Borst, &

Hilburn, 2013). Interaction of **traffic density** with the arguably most important factor, **automation reliability**, is very important in influencing ATCOs' interaction with automation during ATC operations, yet has not received any attention in the current literature.

2.2.3.3 Time Pressure

In ATC, particularly in conflict detection and resolution, time pressure is closely related to Look Ahead Time (LAT). LAT is defined as the maximum amount of time prior to a predicted conflict within which the alerting system is able to clearly indicate that a real conflict is present (Thomas & Wickens, 2005; Thomas, Wickens, & Rantanen, 2003). Thomas, Wickens, and Rantanen (2003) classified LAT into three categories: emergency, tactical, and strategic, which represent immediate action to resolve the conflict within less than 1 minute, consideration of several alternatives in resolving the conflict within 5 minutes, and a flight plan with slight modifications made up to 20 minutes prior to the conflict, respectively. However, due to uncertainties, longer LAT may produce more false alerts than shorter LAT.

Research examining the effect of time pressure on pilots' manoeuvring preferences revealed that under higher time pressure (less time available), pilots showed a higher tendency to use vertical manoeuvre (Thomas & Rantanen, 2006; Thomas & Wickens, 2005). This manoeuvre is preferred because vertical manoeuvre demands less time for execution, lower cognitive effort than lateral manoeuvre, and has a smaller distance in separation standard (Thomas & Wickens, 2005).

In relation to LAT, Eyferth, et al. (2003) conducted a study with ATCOs as the subjects. They recorded the ATCOs' conflict detection rate at the 7 minutes, 5 minutes, and 3 minutes LAT over three conflict classifications (i.e. crossing, descending, and same level). The results demonstrated that more conflicts were detected at 3 minutes LAT than that of 5 and 7 minutes LAT, supporting the uncertainty issue mentioned above. The lowest numbers of conflicts were detected at 7 minutes LAT.

Metzger and Parasuraman (2005) carried out a study in conflict detection and found a different result. They set times of 2 minutes and 4 minutes prior to a conflict after the conflict detection aid failed. No significant differences could be observed in the ATCOs' performances given the 2 minutes and 4 minutes time periods available.

In sum, ATCOs were able to detect conflicts better in shorter LAT (i.e. high time pressure) although there was no significant result in Metzger and Parasuraman's (2005) study.

The factor of time pressure imposed on ATCOs can be a source of worse performance because reaction time may vary with cognitive load. Given smaller intervals between events, reducing the reaction time by dropping feedback may be opted for by the human processor. However, response accuracy may also be influenced (Kantowitz & Sorkin, 1983). Yet, the effect of time pressure in the ATC context, particularly conflict resolution, has not been examined in prior literature.

2.2.3.4 Display

Displays play a very important role in HAI ATC. Relevant information has to be effectively extracted by ATCOs in order to comprehend the present situation and to

determine possible future conditions (Nunes, 2003) for monitoring, conflict detection, and conflict resolution tasks. In ATC, these tasks are supported by automation. However, direct observation of the traffic information and “what automation is doing” is impractical, hence perception of automation-related information is usually mediated by displays (Lee & See, 2004). Two aspects of display have been examined in prior studies in the ATC domain: display dimensionality and cuing.

Display dimensionality in Cockpit Display Traffic Information (CDTI), in particular, has been found to affect pilots’ manoeuvring preferences in avoiding air traffic conflict (Thomas & Rantanen, 2006). As displays provide a graphical representation of vertical dimension, pilots show a greater propensity to apply more vertical manoeuvres. Lateral manoeuvre was found to be more dominantly implemented when pilots used 2D-planar CDTI (Thomas & Wickens, 2005). Examining six different display parameters, Wickens (2000b) provided a summary explicating the costs and benefits of different display dimensionalities. Tethered 3D-display has lower cost in terms of scanning process than that of immersed 3D-display and 2D co-planar display, but it has a higher cost in terms of sight ambiguity (see Wickens, 2000 for detail).

Display cuing can also influence ATCOs’ performance. Imbert, et al. (2014) examined the effects of cue design on attentional costs and failures. Attentional cost was defined as the disruption of tasks due to overly salient design. Attentional failure was defined as missed notifications due to non-salient design. Five cue designs, which were color, color blink, dialog box, shadow mask, and halo, were examined. The results indicated that color and color blink designs led to slower validation times and unnoticed notifications. Halo and shadow mask designs provided attentional power but disrupted the

performance of the primary task. Dialog box achieved a balance between attentional costs and failures.

However, the effects of display cuing may vary depending upon other factors, such as task demands. Metzger and Parasuraman (2006) conducted a study by manipulating display cuing. They studied three kinds of display cuing: normal display, enhanced visual display, and enhanced visual display plus tone. The results showed no significant differences in ATCOs' performances and trust ratings over the three types of display cuing. In contrast, Endsley, Sollenberger, and Stain (1999) examined an enhanced predictive display on a radar screen with enriched aircraft information on the aircraft transition list. The results revealed that the use of a predicted display triggered better ATCOs performances. This indicates that prior evaluation should be carefully conducted when using multimodal cuing for new technology implementation since multimodal cuing may not always bring positive effects; it may also lead to detrimental effects on human performance.

As introduced earlier that direct observation of automation is often impossible, automation transparency has always been closely associated with displays to inform what the automation is doing and why (Norman, 1991; Hoff & Bashir, 2015). Transparency has been found to affect the way users perceive a system and the way users behave toward a system. Automation transparency refers to understandability of the automation that reflects the extent to which an operator is able to comprehend of why and how the automation is doing what it is doing. Transparency has been well studied in many domains. Typically, transparency manipulation was conducted by either giving automation feedback (Bisantz & Seong, 2001; Dzindolet, Peterson, Pomranky, Pierce, &

Beck, 2003), disclosing reliability levels (Wang, Jamieson, & Hollands, 2009), or revealing shared goals between humans and automation (Nass, Fogg, & Moon, 1996; Verberne, Ham, & Midden, 2012).

In the ATC domain, a study regarding transparency was carried out by classifying participants into two groups: a situation-specific reliability (SST) group and a non-SST group (Masalonis & Parasuraman, 2003). Situation-specific reliability was defined as disclosing the reliability of automation where it did not consider deviations from the flight plan, thus the reliability would suffer during the scenario experiment. The result showed that transparency did not affect performance but activated the tendency to report the system's signal liberally.

In command and control context, Rovira, et al. (2014) conducted a study to examine transparency by providing contextual information to a human operator while working with an automated aid. The contextual information in this study referred to the state of factors that affect the reliability automation's recommendation. The results revealed that providing contextual information could offset the performance decrement with imperfect decision automation.

However, the amount of information provided to increase automation transparency should be considered carefully, because it may create too much clutter. In the study of Marusich, et al. (2016), the volume of task-relevant information was varied, and the results showed that task performance was not improved with increased task-relevant information, although it is accurate. Furthermore, it led to poorer SA. Therefore, knowing what information to be displayed and how much it should be displayed is vital. To

increase automation transparency, designers should carefully develop intuitive displays that show dynamic processes that can improve SA (Sebok & Wickens, 2016)

In sum, automation transparency can enhance performance and decrease performance loss resulting from automation failures, although increasing transparency does not necessarily increase performance. In the current experiment, transparency will be implemented with a vertical situation display described below that provides the ATCOs with a better visual explanation for the spatial reasoning behind the CRA advice.

2.2.3.5 Vertical Situation Display

This section reviews studies pertaining to vertical situation display covering two primary features: display dimensionality (i.e. 2D co-planar and 3D display) and context of tasks (i.e. ATC traffic tasks and pilot traffic tasks) as shown in Table 2.3. The distinction between the left and right columns is obvious in determining study relevance. The distinction between the top and bottom rows is that when vertical information is integrated into a 3D display that information becomes perceptually ambiguous (Wickens, 2000). It should also be apparent that the cell in the upper left is most critical, defining the VSD used in the current research. The cell in the lower right is least similar to the current research. Hence studies here will be described in less detail than those in the two remaining cells. Vertical information is vital since data from the ASRS (Aviation Safety Reporting System) disclosed that about 63% out of 300 accidents were related vertical due to lack of vertical understanding (Vakil et al., 1996).

Table 2. 3. Relevant studies of VSD

Display dimensionality	Context of tasks	
	ATC traffic tasks	Pilot traffic tasks
Vertical plan view	N.A.	<ul style="list-style-type: none"> - Prevot & Palmer, 2000 - Vakil, et al. 1996 - Heylen, et al. 2008 - McElreath & Pingsterhaus, 2000 - Merwin & Wickens, 1996 - Alexander, Wickens, & Merwin, 2005 - Wickens, et al. 1997
Vertical integrated into 3D view	<ul style="list-style-type: none"> - Wickens, Miller & Tham, 1996 - Tham & Wickens, 1993 	<ul style="list-style-type: none"> - Merwin & Wickens, 1996 - Wickens, et al. 1997 - Alexander, Wickens, & Merwin, 2005 - Wickens, et al. 1996 - Wickens & May, 1994 - Nguyen, et al. 2011

a. Studies related to pilot tasks with vertical plan view

Pilots on board are currently supported with the vertical situation display – that is, using continuous graphics to present vertical information -- that provides vertical profile along the aircraft's path in the form of a side-view vertical profile display (i.e. co-planar display).

Prevot and Palmer (2000) tested the concept of VSD in a full mission simulation study (i.e. with and without the VSD). On the VSD (Figure 2.6 and Figure 2.7), the active

path, the pre-programmed Flight Management System (FMS) path, the flight extrapolation path, the aircraft, and the waypoints are represented by the white line, the purple line, the green line, the white single triangle, and the white twin triangles, respectively. The VSD tested had two different modes (i.e. full-mode and co-planar mode) that could be changed by the pilot depending upon their preferences. The full-mode (Figure 2.6) represented only the VSD while being selected. The co-planar mode (Figure 2.7) showed both planar display and VSD on the same screen at the same time.

Figure 2. 6 Full-mode VSD (Prevot and Plamer, 2000)

Figure 2. 7 Co-planar VSD (Prevot and Palmer, 2000)

Prevot & Palmer (2000) conducted a full mission simulation study to improve the ground-flight deck automations compatibility. They specifically investigated cockpit interfaces that could support this aim. They integrated a Missed Approach Point (MAP) display and a VSD to support the flight crews' understanding of the navigation task, aircraft position, and vertical velocity (Prevot, 1998). In the study, the VSD was set as an independent variable with two levels: shared co-planar VSD mode and full VSD mode. Subjectively, by having the VSD available, the majority of crew members felt "more ahead" of the airplane. When the VSD use was opted for, shared co-planar mode was clearly preferred and used almost all the time (>98%). However, Prevot and Palmer (2000) did not find any significant effect of the VSD on overall crew performance and

workload. This might be due to less challenging scenarios wherein traffic density and weather conditions were not high and worse, respectively.

Vakil, Midkiff, and Hansman (1996) designed an electronic vertical situation display (EVSD). The EVSD shown in Figure 2.8 has more complete features than the general VSD. Vakil, et al. (1996) incorporated the anticipated modes and mode transition locations into the EVSD. This provides pilots with further specific actions. The line colours represent similar features as the general VSD. These colour symbols are based on the colour agreement used in aviation display design.

Vakil, et al. (1996) evaluated an electronic vertical situation display (EVSD) and at least half of the participants subjectively rated the EVSD as very valuable. This might be because the EVSD (Vakil, et al., 1996) provided more specific information on how to act (e.g. low speed as an anticipated mode) given the corresponding vertical situation. The results of the EVSD use were promising as the anticipated mode of the EVSD allowed pilots to predict the under-speed situation 60.6 seconds sooner with than without the EVSD, at a confidence level of 95%. In addition, although little statistical significance was found, subjects were also observed to deviate from the target altitude by 45.8 feet less with than without the EVSD.

Figure 2. 8 Electronic Vertical situation display (EVSD) (Vakil, et al, 1996)

Vertical Separation Assistance Display (VASD) was also designed to provide more comprehensive information regarding the vertical profile and aircraft capability in climbing and descending (Heylen, Van Dam, Mulder, & van Paassen, 2008). Heylen, et al. (2008) compared the effect of the VSD and the VSAD on Situation Awareness (SA) and found that SA was much higher with the VSAD. Another study (Suijkerbuijk, Borst, Mulder, & Van Paassen, 2005) manipulated the display into three categories: (i) a Synthetic Vision System (SVS) with a Horizontal Situation Display (HSD), (ii) the SVS with the HSD and the VSD, and (iii) the SVS with the HSD, the VSD, and the aircraft performance information. The results indicated that the VSD could successfully assist pilots in maintaining their situation awareness.

Vertical situation display (VSD) has been found to affect pilots' situation awareness, workload, and performance in several ways. McElreath & Pingsterhaus (2000) suggested that providing vertical situation information comprised of some sources in an arranged form will allow pilots to accurately assess a complete vertical situation using minimum cognitive processing, thus inducing lower mental workload.

Merwin & Wickens (1996) evaluated two perspectives and a coplanar display format for conflict avoidance in a free-flight environment. Similar to previously reported results, the coplanar display was found to support performance and flight safety under all of the experimental conditions, as indicated by fewer extreme maneuvers and fewer occurrences of actual conflicts.

Alexander, Wickens, & Merwin (2005) conducted three experiments to compare the 2D coplanar display and the 3D display. The first experiment was carried out with less traffic, and both display formats exhibited similar effects on performance. Comparable findings were also found in the second and the third experiments with higher traffic. It was surprising that performance using the coplanar display was somewhat disrupted as traffic density increased. However, the use of coplanar displays could support a slightly lower workload across all three experiments.

Wickens, et al. (1997) examined display dimensionality (2D coplanar vs 3D) by imposing a weather hazard. The results showed that coplanar display better supported flight safety measures in terms of actual traffic conflicts, predicted traffic conflicts, and weather conflicts. In addition, performance loss was observed in actual conflicts with the perspective displays, especially in the non-level conflict geometries (i.e. climbing or descending).

In sum, the 2-D co-planar display, explicitly portraying the altitude axis in an unambiguous, non-compressed format, was generally found to be useful in assisting pilots to safely avoid predicted traffic conflicts in comparison to a 3D perspective display due to the ambiguity of the latter (Alexander, Wickens, & Merwin, 2005; Thomas & Wickens, 2006; Wickens, 2000).

b. Studies related to ATC traffic tasks with vertical information integrated into 3D view

Two studies were conducted to examine the use of planar (the existing plan view radar display that only shows lateral and longitudinal planes) and perspective display formats in ATC tasks and the effects on ATCOs' performances (Tham & Wickens, 1993; Wickens, Miller & Tham, 1996).

Wickens, Miller & Tham (1996) conducted a study to examine the effect of display dimensionality (3D vs 2D) on ATCOs' performance in safely granting pilots' manoeuvre requests. The results indicated that no cost of the 3D display was observed for the accuracy measure. However, ATCOs responded to the requests faster when given the 2D display than when given the 3D display.

Tham and Wickens (1993) carried out two phases of study to compare the use of a perspective ATC display to the use of a plan view display on ATCOs' performances. In Tham and Wickens (1993), common ATC tasks and a traffic management task were performed in Phase 1 and Phase 2, respectively. The results showed that the plan view display benefitted decisions judged along the planar axes, although no difference was evident between the two displays in the conflict detection task. Similar efficiency between

the two display formats was also observed in the traffic management task. However, slower detection of uncertain manoeuvres deviation was observed when using the perspective display.

From the ATCOs' point of view, it could be inferred that the existing plan view radar display could support more precise estimation and faster judgment as compared to 3D perspective display, because of the biasing/ambiguity problems depicting the altitude axis in the latter. This might be due to different tasks' requirements, in which ATCOs are trained to use the 2D display with fixed viewpoints that can help them accurately estimate critical distances by monitoring the environment from an exocentric view (Azuma, Daily, & Krozel, 1996). Although there are no differences in the traffic flow efficiency, perspective display (involving vertical projection) tended to show a greater cost while dealing with unexpected changes in lateral manoeuvres (i.e. heading and airspeed).

c. Studies related to pilot traffic tasks with vertical information integrated into 3D view

In the studies by Merwin & Wickens, 1996 and by Wickens, et al. 1997, the 3D perspective display was found to have costs in flight safety in terms of traffic and weather-related conflicts. Significant performance loss when traffic was non-level was also observed when using the 3D display. In addition, it led to higher pilots' workload (Alexander, Wickens, & Merwin 2005).

However, Wickens, et al. (1996) found that pilots responded to both 2D coplanar display and 3D display equally in speed. This result corresponds to the results of Wickens and May (1994), who found better pilot performances after some experience in using a perspective display. Another study (Nguyen et al., 2011) concerning the impact of data-

communication revealed no significant difference between the two display (2D vs 3D) formats on the pilots' workload as well. However, there was evidence of the 3D display contributing to an increase in the pilots' understanding of the programmed flight plan when the non-integrated data-communication method was used.

In a word, although there were inconsistent findings with regards to the benefits and costs of a 3D display for pilots, however the 3D display has more costs than its offered benefit.

Collectively, from the above-mentioned studies, it can be concluded VSD has been found as a beneficial support to increase pilots' performance and SA. In addition, it was clear that the 2D plan view display is preferred over the 3D one (the perspective display). The performance measures also indicated that the 2D display benefits ATCOs and pilots more than the 3D display. The reasons of these benefits of the 2D plan view over the 3D display lie on the issues of display ambiguity and display clutter. However, it is worth to note that there is **no study that has examined the vertical plan view (co-planar) for ATC tasks**. However, the predicted vertical state for conflicting aircraft reflects automation's belief in what the aircraft will be doing at the moment the loss of separation occurs (Mercer, Gomez, Homola, & Prevot, 2014). Since the benefits of the plan view VSD in supporting pilots are clear, the concept of plan view VSD for ATC facilities deserves an investigation as an attempt to deal with the increasing air traffic. In the case of evaluating the CRA automation, the benefits of such a display transparency may be particularly realized in the context of supporting performance during automation imperfections

2.2.3.6 Complexity

Several studies have attempted to define complexity and to specify factors affecting the construct in ATC (Kopardekar & Magyarits, 2002; Mogford, Guttman, Morrow, & Kopardekar, 1995; Pawlak, Brinton, Crouch, & Lancaster, 1996; Sridhar, Sheth, & Grabbe, 1998; Xing & Manning, 2005).

Mogford, et al. (1995) classified many complexity factors from prior studies into two main sources of factors affecting ATC complexity: air traffic pattern and sector characteristics. Sridhar, et al. (1998) defined complexity as “airspace complexity” instead of ATC complexity which depends upon structural characteristics and the airspace flow.

Pawlak, et al. (1996, p.6) identified 16 different factors affecting ATC complexity. To measure ATC complexity, a method referred to as Dynamic Density (DD) has been introduced (Kopardekar and Magyarits, 2002) and DD metrics better represent complexity than aircraft count alone. A unified DD metric consisting of all variables from the four proposed metrics could provide the best results in all conditions involving different facilities.

Another form of complexity was proposed by Boag, Neal, Loft and Halford (2006), that is relational complexity that was seen from the point of view of aircraft transition and separation variables.

Histon, et al. (2002) classified many factors influencing cognitive complexity as reported by ATCOs into three major categories: airspace factors, traffic factors, and operational constraints. This classification seems comprehensive since it covers most factors that were explicated in other studies concerning complexity.

2.2.3.7 ATCOs' Mental Model

A Mental Model (MM) is defined as “a mental structure that reflects the user’s understanding of a system” (Carroll, Olson, & Anderson, 1987 in Wickens, et al. 2013 p.236) and thus it involves expectations of a system’s response (Wickens, et al., 2013). The mental model is also defined as the system knowledge and is part of Long Term Memory (LTM) (Kallus et al., 1999) thus, can be prearranged into LTM through training (Wickens, et al. 2013).

In en-route ATC, the majority of ATCOs initially check the radar and then check the Flight Progress Strips (FPSs) and other sources of required information. ATCOs’ actions in checking and integrating the actual information into MM is required to set their Mental Picture (MP) of current traffic conditions and the initial inferences of previous en route ATCOs’ sector plans (Kallus et al., 1999). Niessen, et al. (1999) developed a model of ATCOs’ pictures, comprised of five modules: data selection, anticipation, conflict resolution, update, and control.

ATCOs’ ability to effectively and safely maintain aircraft separation at all points in the airspace depends upon maintenance of their MP (Nunes & Mogford, 2003). To do this, ATCOs must continually perform correct mental computations, including updates for speed changes, altitudes, headings, weather, and the operational implications of such changes (Murphy et al., 2012).

2.2.3.8 Conflict Geometry

The definition of conflict geometry involves factors as follows: (i) lateral angle formed by the intersecting trajectories between two conflicting aircraft, (ii) airspeeds, (iii)

altitudes and dynamic changes involving combinations of the factors (Thomas & Wickens, 2005). The schematic representation of conflict geometry is depicted in Figure 2.9. Ownship represents aircraft, which becomes the point of reference, and intruder represents conflicting aircraft.

Figure 2. 9 Schematic Representation of Conflict Geometry

As for ATCOs' point of view, Loft, Bolland, and Humphreys (2007) in Fothergill, Loft, and Neal (2009) highlighted that increases in minimum lateral separation distance induced ATCOs to intervene more actively.

In their conflict detection research, Eyferth, Niessen, and Spaeth (2003) investigated three different conflict classifications: crossing, descending, and same level. Similar results as Rantanen and Nunes' (2005) were obtained in that conflict detection involving vertically manoeuvring aircraft was faster and higher confidence levels was observed in ATCOs in comparison to other conflict classification (i.e. crossing and same level).

2.2.3.9 Expertise

The level of expertise can be typically classified as how long a person has been working as an ATCO and how much experience the person has. While expertise is assumed to affect the performance of ATCOs, the effects of this construct actually vary. Loft, et al. (2007) in Fothergill, et al. (2009) investigated the effect of ATCOs' experience on the intervention decision. The results indicated that the experts did more interventions while controlling aircrafts in airspace than the trainees.

In contrast, Rantanen and Nunes (2005) found no difference in performance patterns for both expert and novice groups. Similarly, Niessen, et al. (1999) found no significant differences among experienced, intermediate, and less-experienced ATCOs in accessing radar information and flight-strip information. Also no relevant difference was observed in classifying airspace control situations for both experienced and less-experienced ATCOs.

Surprisingly, Borst, Westin, and Hilburn (2012) found that students outperformed ATCOs performance under low and high complexity. Students appeared to interact less with traffic than ATCOs in all parameters. Students were found to have higher control efficiency than ATCOs. However, maneuver preferences between students and controllers were similar.

In summary, Section 2.2.3 reviews studies pertaining to the drivers of human-automation interaction in ATC context. Strong empirical evidence regarding the effects of automation reliability and time pressure has been found in conflict detection and aviation-related research. However, the automation reliability and time pressure has never been

studied in conflict resolution, the process that actually involves more problematic issues in its development and application due to different value structure and possibly greater costs when the automation fails (Sebok & Wickens, 2016).

Furthermore, the reliability effects of conflict resolution automation has become increasingly relevant given the increasing traffic, in the sense that the costs and benefits of the automation may be amplified in the high traffic load environment but no study has been conducted to examine this crucial issue.

Next, despite the benefits of 2D planar vertical display, its application has never been examined in ATC facilities. The vertical plan view display is however anticipated as a mean to increase transparency in ATCOs- automation interaction, thus possibly compensates automation's failure.

To fill the research gaps, Chapter 3 will overview three experiments examining the four factors on ATCOs performance variables.

2.3 Conflict Detection and Conflict Resolution System

2.3.1 Support Tools for Conflict Detection and Conflict Resolution

There are several automation tools that are used for conflict detection and resolution in the ATC systems as shown in Table 2.4. The discussion regarding each automation tools is provided as follows:

Table 2. 4. Automation supports for conflict detection and resolution

	Conflict Detection	Conflict Resolution
In ATC facilities	<ul style="list-style-type: none"> - Short term conflict alert (STCA) - User request evaluation tool (URET) 	N.A
On-board aircraft	<ul style="list-style-type: none"> - Cockpit display of traffic information (CDTI) (not available yet in commercial aircraft) 	<ul style="list-style-type: none"> - Traffic alert and collision avoidance system (TCAS)

In the current ATC facilities, ATCOs are equipped with conflict detection tools that support short term and long term conflict detection. The short term conflict alert (STCA) is an automation tool that works within the tactical range of time to detect and to alert ATCOs of an impending conflict (Wickens, Rice, et al. 2009). This tool is embedded in the current radar display and warns ATCOs of the predicted conflict around 90 seconds prior to collision (Thomas & Wickens, 2006). The conflict detection in ATC facilities is also supported by the User Request Evaluation Tool (URET), which provides controllers with a probe that can predict a collision 20 minutes in advance (Brudnicki & McFarland, 1997). The main feature of URET is to allow D-side ATCOs to try planning several possible trajectories and identifying how they will affect the whole flight plan (or so-called what-if scenario) (Wickens, 1998). However, since this tool provides long-range planning of up to 20 minutes, the uncertainty of trajectory planning is also present due to potential sudden weather changes or other flight deviations.

For on-board aircraft, conflict resolution in the flight deck is supported by the Traffic Alert and Collision Avoidance System (TCAS). TCAS is an airborne automated system that provides visual advice to pilots if there is another aircraft within vicinity (i.e. where ATCOs operational error exists), provides alert visually and aurally for an imminent possible collision (i.e. when both aircraft are still on the same course), and issues a resolution advisory in both visual and auditory forms (i.e. auditory advice such as “climb climb climb” (FAA, 1990)), that instructs vertical manoeuvre for pilots to avoid the possible collision (Wickens, et al. 1998) . There are two types of TCAS that are available in the current operations. TCAS I provides pilots with traffic advisories of possible dangerous intruder (TAs). TCAS II, in addition, provides resolution advisories (RAs) of vertical evasive manoeuvre (Rantanen, et al. 2004). That is, like the CRA advice for ATCOs. The TCAS is activated around 30 to 45 seconds before the collision (Thomas, Wickens & Rantanen, 2003; Rantanen, et al. 2004) which includes the estimation of pilot response, vertical manoeuvre time, time necessary to achieve a climb or a descent rate, and transition time.

Another airborne automation support that can be used to help pilots understand the surrounding airspace and the traffic conditions is the Cockpit Displays of Traffic Information (CDTI), however, it is not yet available in commercial transport. The CDTI receives information about current aircraft position, heading and velocity, and information about flight plans, which are obtained from the Automatic Dependent Surveillance-Broadcast (ADS-B) and the Traffic Information Service (TIS) (Thomas & Wickens, 2006). The function of CDTI is very different from the TCAS. TCAS is the “last line of defence” applied in an emergency situation timeline (Rantanen, et al. 2004; Thomas &

Wickens, 2005; Thomas & Wickens, 2006). CDTI is designed as a more strategic planning tool to support pilots' understanding of air traffic in the airspace and to maintain self-separation. The CDTI also assists pilots in creating new flight paths if there is a predicted conflict. The CDTI can be fully reliable if it is only used to present current airspace situation. However, when the display function of the CDTI is coupled with an advisor used to predict conflicts and the resolution manoeuvres, it may be imperfect due to uncertainty of the future behaviour of all aircraft.

From the review of the existing automation aids used both in ATC facilities and on airborne aircraft, it can be inferred that automation aids assisting ATCOs in relation to air traffic conflict mainly provide support in detecting possible future conflicts. It is true that the TCAS also provide resolution advice for the pilot. However, the TCAS advice comes less than 45 seconds before the collision (Thomas, et al., 2003), and is entirely pilot-focused. In addition, ATCOs are often frustrated because there is no feedback for ATCOs when the TCAS is activated so they have no idea when an aircraft suddenly ascended or descended because the TCAS alert might have gone off. In such a limited time, it is a very urgent situation where misunderstanding between the pilot and the air traffic controller is non-permissible.

The projection of the extensive increase in air traffic density creates the plausible possibility of more traffic conflicts. Facing this tight situation, automation supports providing conflict probes and conflict alerts will no longer be sufficient. Hence, the concept of automated conflict resolution, particularly for ATCOs in ATC facilities, which is currently not available, becomes more relevant and important; this is described in more detail below.

2.3.2 Operational ATC Concepts for Conflict Resolution in ATC Facilities

2.3.2.1 ATCOs Manually Resolve Conflicts

The process of solving conflicts has been well elaborated upon by Kallus, et al. (1999). The comprehensive task process of solving conflicts is illustrated in Figure 2.10.

The conflict solving process is initially started when a predicted conflict exists. If it requires immediate action from en route ATCOs to resolve it, ATCOs will open their “conflict solution library” in their long term memory; if no immediate action is needed (perhaps more time is available before the actual conflict), ATCOs will monitor the aircraft situation. When ATCOs retrieve a routine solution from their memory, they will evaluate the solution and match it with the conflict condition. If ATCOs perceive that the solution is optimal given the available time, then they will coordinate with pilots to issue instructions and the manoeuvring result will be the next input in their mental picture (MP). If the designated solution is not the best solution, then ATCOs will need to generate a new solution, which may be difficult for them, especially in a high traffic density environment. If a conflict involves more than two aircraft, a negotiation process is required as a part of the coordination process (Rong, Geng, Valasek, & Ioerger, 2002).

Figure 2. 10 Solving Conflict Process (Kallus, et al. 1999)

The conflict resolution process has not been extensively studied, especially from the point of view of ATCOs. This is important since although the Federal Aviation Administration (FAA) has provided definitions and implementation guidelines for each

type of separation standard, no such guidelines have been provided on how to resolve various conflicts in an effective and expeditious manner (Rantanen & Wickens, 2012). No formal instructions are given to ATCOs to provide solutions to conflicting aircraft (Thomas & Rantanen, 2006). There are few studies have tried to address this issue.

Rantanen and Nunes (2005) conducted a study examining the effects of all conflict geometry parameters on conflict detection performance. They hypothesized that hierarchical strategy in conflict detection exists based on the order of time cost and required mental effort. The results demonstrated that altitude was the initial information being examined by ATCOs to determine the existence of potential conflict. ATCOs also did not make no-conflict decisions based on the two diverging aircraft. For the converging aircraft, ATCOs' responses were faster when the monitored aircraft were flying at the same speed than at a different speed. In the opposite headings condition, an interpretation that the ATCOs initially verified the aircraft's altitudes was also observed. To conclude, ATCOs were found to apply hierarchical strategy in conflict detection decisions by first checking altitude information; if the monitored aircraft were vertically separated, they made a no-conflict decision without considering other traffic characteristics, such as conflict lateral angle or aircraft speed.

Loft, et al. (2007) provided a synthesis of literatures examining ATCOs' strategies for conflict resolution. Seven studies were examined and all of the studies used interviews to gain information from ATCOs about their strategies. Although there are no formal guidelines in resolving conflicts, the majority of the studies revealed that there is a pattern to conflict resolution strategies similar to Rantanen and Nunes (2005) found, that most

ATCOs indicated that they would initially examine the altitudes of conflicting aircraft and then they would focus on critical points where conflicts have previously occurred.

Although no formal guidelines are available, ATCOs typically gain conflict resolving techniques during their initial training and through their own experiences. These then cause ATCOs to form a “CR library” in their LTM. Rantanen and Wickens, (2012) performed an analysis to investigate real data from five different ATC en route centers in the United States. “Regular patterns” in selecting conflict resolution manoeuvres were observed in objective data through objective analysis of 495 cases of conflict alert. This was the first analysis conducted using real data to empirically demonstrate the existence of “ATCOs’ CR library”.

The observed pattern was that ATCOs preferred vertical manoeuvres over lateral manoeuvres (i.e. turns). The pattern indicated that ATCOs tended to initially issue vertical manoeuvres unless it was impossible. Subsequently, ATCOs were more likely to issue lateral manoeuvres unless it was impossible. Afterward, ATCOs would probably command for speed manoeuvres unless it was impossible and so forth (based on the expected utility in Rantanen and Wickens, 2012). This expected utility was supported by the study in Mercer, et al. (2014) that shows that altitude manoeuvres were preferred to resolve over 61% of the conflicts while lateral manoeuvres were preferred to resolve 31% of the conflicts, and only the remaining 8% of the conflicts were resolved by speed or combination of the manoeuvres.

Workload level has also been seen as a variable affecting ATCOs’ manual conflict resolution process. Fothergill and Neal (2008) in Fothergill, Loft and Neal (2009) conducted a study examining workload level on conflict resolution strategy. The results

indicated that when an optimal solution was difficult to obtain, ATCOs were less likely to select the optimal solution under high workloads than under low workloads. Alternatively, when the conflict was categorized into simple conflicts, ATCOs would use an optimal solution to resolve it under both high and low workload conditions.

The manual conflict resolution process will be more challenging for ATCOs in the future ATM system due to the increase in traffic density, which will likely trigger more air traffic conflicts, and of course, smaller manoeuvring space. Hence, ATCOs should be supported by the automation in performing their tasks. The concept of automated conflict resolution will be elaborated upon in more detail in the following section

2.3.2.2 Automated Conflict Resolution Concepts

The concept of automated conflict resolution has been explored in several studies (Table 2.5). Generally, it has been classified into two types of CRA: fully automated CRA and interactive CRA. The fully automated CRA refers to conflict resolution (CR) tasks that will be performed fully by automation without ATCOs and pilot intervention. The interactive CRA reflects the function allocation between humans (ATCOs and pilot) and automation in solving conflicts, with automation providing advice that may be rejected by the ATCO before giving instructions to the pilot.

Table 2. 5. Studies related to automation conflict resolution

Study name	Proposed CRA algorithm	Evaluated CRA algorithm in a HITL	Compared CRA algorithm to unassisted ATCO resolution
Erzberger (2006)	Resolution trajectories are	No	NA

	generated by evaluating successive alternative trial manoeuvres (Resolution Aircraft and Manoeuvre Selector (RAMS)).		
Sanchez (2009)	NextGen Controller-Automation Interaction Display (N-CAID) gives ATCOs three options: 1) accept, 2) request new resolution, 3) modify the generated resolution	No	NA
Erzberger & Heere (2010)	Tactical Separation Assured Flight Environment (TSAFE) gives resolution advice on aircraft's data-tags	No	NA
Cardosi, et al. (1992)	Resolution advisories are given in the form of manoeuvring abbreviations	Yes-ATC	Did not compare with manual.
Dwyer & Landry (2009)	"Resolution options" provides modifiable automated resolutions. "Auto resolutions" provides fully automated resolution sent to FMS.	Yes-ATC	CRA helped
Prevot, Homola & Mercer (2008)	Interactive CRA mode provides modifiable on-demand resolutions. Fully automated CRA mode: the automation does all CR functions.	Yes-ATC	CRA helped
Prevot, et al. (2009)	TSAFE (examined Erzberger & Heere's algorithm)	Yes-ATC	CRA helped
Jha, et al. (2011)	Integrated conflict detection,	Yes-ATC	CRA hurt

	negotiation information, and conflict resolution display		
Borst, Westin, Hilburn (2012)	Solution Space Diagram (SSD): a tactical decision-support tool vectoring aircraft in conflict-free manoeuvres	Yes-ATC	CRA helped
Westin, Borst, Hilburn (2013)	SSD	Yes-ATC	CRA helped
Cabrall, et al. (2014)	Automated trajectory-based CR for more than 3 mins prior to LOS and TSAFE for less time before LOS.	Yes-ATC	CRA helped
Gorithzlehner, et al. (2014)	SSD	Yes-ATC	CRA helped
DiMeo, et al. (2001)	Conflict resolution tool embedded in a Cockpit Display of Traffic Information (CDTI) for pilots	Yes-CDTI	CRA helped
Canton (2005)	CRA embedded in a CDTI for pilots	Yes-CDTI	CRA helped
Dao, et al. (2011)	CRA embedded in a CDTI for pilots	Yes-CDTI	Not sig.
Vu, et al. (2012)	CRA embedded in a CDTI for pilots	Yes-CDTI	CRA hurt

A concept of fully automated CR has been evaluated (Erzberger, 2006). The fully automated CR in this study was defined as resolution trajectories generation that could be sent to the flight deck automation in airborne aircraft via a data-link. The resolution trajectories were generated by evaluating trial resolution manoeuvre alternatives. This iteration was run continuously until it reached a safe and successful resolution (see Figure

2.11 for example). Erzberger (2006) then conducted a traffic simulation study examining this concept and found that it is able to resolve conflict more efficiently at higher traffic densities, but it was not examined with a human-in-the-loop (HITL) study.

Figure 2. 11 Example of Resolution Trajectory (Erzberger, 2006)

Sanchez, et al. (2009) developed a prototype called NextGen Controller-Automation Interaction Display (N-CAID). The N-CAID has a feature to make transparent automation-generated resolutions of the CRA. N-CAID leaves ATCOs with three options in solving conflicts. If ATCOs agree with the automation-generated resolution, they can press the accept button and the automation will execute the solution through Data-com to the respective aircraft. If ATCOs reject the automated resolution, they have two choices. ATCOs can request the automation to calculate and display a new resolution maneuver. On the other hand, ATCOs can also slightly modify the automation-generated resolution. No simulation testing was done using the N-CAID, hence the benefits of N-CAID have not been explored.

Erzberger and Heere (2010) proposed a concept for tactical conflict resolution. The concept was named Tactical Separation Assured Flight Environment (TSAFE). TSAFE would initially alert ATCOs by displaying a message containing the data of two aircraft in conflict, including the resolution advice on the aircraft's data tags two minutes prior to the conflict. ATCOs would then be given 1 minute to choose one of following actions: prevent TSAFE from issuing a resolution to the aircraft, command TSAFE to send the resolution to the aircraft, or take no action at the given time. If no resolution is made within that one minute, TSAFE would issue an immediate resolution advisory while simultaneously informing the ATCOs of the executed actions. However, ATCOs and pilots must understand TSAFE's roles and conditions when advisories are issued. This will trigger complex human-factor issues which must be further investigated using human-in-the-loop simulations (Erzberger & Heere, 2010).

The concept of CRA has been researched since more than two decades ago (Cardosi et al., 1992) yet has not been implemented today. In 1992, Cardosi, et al. examined an automated tool for ATCOs at en-route facilities. The CRA performed computation and validation for the predicted conflict based on the conflict alert and displayed a single resolution for ATCOs. The CRA displayed the advice for ATCOs on resolving a conflict between aircrafts in the form of abbreviations for each aircraft like R for right and L for left. The ATCOs' response time and acceptance ratings were obtained when using the CRA and they did not compare the performance with manual conflict resolution. The average response time required for ATCOs to read, to understand the CR advice, and to decide on the resolution was 18 seconds. Regarding the CRA acceptance, majority of ATCOs indicated that they were reluctant to use the CRA and they would at

least like to use it as a backup in solving conflicts. This result was arguably because the traffic in the last two decades was still much more manageable than today.

However, the concept of CRA has become increasingly relevant with the increasing traffic and has received more attention in a number of more recent previous studies. A previous study (Dwyer & Landry, 2009) provided a review on collision avoidance and conflict resolution concepts. Three main concepts were examined: (i) Conflict Detection with Resolution Tools, (ii) Conflict Detection with Resolution Options, and (iii) Conflict Detection with Auto Resolutions. In the first concept, ATCOs were provided with decision support tools for scheduling and trajectory planning. The simulation results of this concept indicated that capacity improvement would not be obtained without the ability to directly send trajectory updates to the flight deck automation. In the second concept, ATCOs were given an automated conflict identification and resolution option. These features could be displayed whenever required by the ATCOs and could also be uplinked once accepted or modified by the ATCOs. Using this concept, ATCOs' workloads were greatly reduced and a 300% increase in traffic was still seen as manageable. The third concept was fully automated; this means that automation electronically sent a resolution to the aircraft and the resolution was then be implemented by the pilot or the flight management system (FMS). This concept puts ATCOs out of the loop. Similar to the second concept, ATCOs' workloads were reduced and this allowed a 300% traffic increase to be managed.

Prevot, Homola, & Mercer (2008) compared three CR concepts in their study: manual (ATCOs solve conflicts manually), interactive mode (ATCOs were given automated resolution advice upon request; ATCOs may also modify or reject the

resolution), and fully automated mode (no ATCOs intervention required for solving conflicts; automation generates a solution and sends it directly to the airborne aircraft for maneuvering implementation). The results revealed that both fully automation and interaction mode supported ATCOs performance as indicated by less delay in conflict resolution as compared to manually performing the task. In addition, significantly higher numbers of conflicts and higher workload were also found when ATCOs worked manually. Interestingly, the results revealed a better performance trend in the interactive mode than in other modes. This result suggests that ATCOs working together with automation would produce more effective performances than either alone. This finding also corresponds to Metzger and Parasuraman's (2001) result that ATCOs performed better while actively being involved in active monitoring than passive monitoring.

Prevot, et al. (2009) was the first HITL study that examined the TSAFE concept (Erzberger & Heere, 2010). The study varied two conditions (TSAFE vs No-TSAFE) across two traffic densities (medium vs high traffic loads). ATCOs were found to have higher successful conflict resolution rates with TSAFE than without it, in both traffic densities. However, the workload was not significantly different between TSAFE and No-TSAFE conditions. ATCOs rated the TSAFE condition as more acceptable across traffic densities than No-TSAFE condition. Overall, the results showed benefits of the TSAFE as a tactical safety layer in ATC in terms of safety and concept acceptability.

Implementing the concept of negotiation between conflicting aircraft, Jha, et al. (2011) investigated an automation aid that integrated conflict detection, negotiation information, and conflict resolution display. Negotiation information display shows negotiation messages between aircraft in conflict to decide how to effectively resolve the

conflict. The CR display exhibits the automation-generated resolution for the conflict in the form of a bar chart indicating four attributes: altitude high, altitude low, heading right, heading left. A HITL simulation study was also conducted using the designed CRA. It was surprising that the presence of the conflict resolution display did not positively contribute to a better performance. When more free-flying aircraft were available, the ATCOs' performance became even worse. This might be due to the verification process required by ATCOs prior to applying the resolution. ATCOs also stated that they would like to rely more on the radar screen than the automation aid.

Borst, Westin and Hilburn (2012) examined a conflict detection and resolution tool namely the Solution Space Diagram (SSD) that represents an aircraft's speed and heading control space. An ATCO can utilize the SSD to vector aircraft into conflict-free areas in order to maintain separations between aircraft. In this study, they compared ATCO and student participants across low and high traffic complexities. They found that students outperformed ATCOs across the two traffic conditions. In all scenarios, students had no LOS whereas ATCOs created LOS. Students were found to have higher control efficiency than ATCOs. ATCOs and students showed different attitudes toward the conflict warnings. Students reacted earlier than ATCOs to the SSD and immediately solved the conflict while ATCOs appeared to disregard the SSD. This might be because ATCOs trusted the SSD less than the students did.

Westin, Borst and Hilburn (2013) manipulated resolution strategy into two conditions. The first condition was conformance strategy, in which ATCOs were provided with their own resolutions, which were treated as automation resolution. Only sector configuration was rotated so that ATCOs did not recognize that the advice was actually

their own. The second condition was non-conformance strategy, in which the ATCO participants were given their ATCO friends' strategies, which were treated as automation-generated resolutions. The results revealed that the CRA was accepted more and responded to more rapidly under the conformance condition than the non-conformance one. The average ATCOs response time was also faster when provided with the conformance strategy.

Cabrall, et al. (2014) conducted a study to evaluate the concepts of an integrated automated trajectory-based conflict resolutions (TBOCR) and TSAFE. The TBOCR generated for resolutions for conflicts more than 3 minutes prior to LOS and TSAFE was used for conflicts with lesser time to LOS. The full study was conducted to simulate four different sectors with varying range of traffic densities within each sector that was set to be double of the current air traffic. The numbers of conflict occasion are different among the sectors. However, although ATCOs worked in a double-traffic level than the current traffic, they rated their workload to be considerably low. In general, safety and capacity could still be maintained with automation, despite the increase in traffic density (Cabral, et al. 2014).

Goritzlehner, et al. (2014) also examined SSD but in this study, they varied the automation's transparency across low and high traffic complexities. The automation's transparency included three conditions: i) T1 or manual (no SSD), ii) T2 (SSD support, only heading), iii) T3 (SSD support, heading and speed). There was a trend of lower SSD acceptability as the traffic increased. No significant effect of transparency was found although there was a trend of decreasing acceptance as transparency increased. However, the performance was better with T3 than with T2 or T1.

For the studies examining CRA embedded in CDTI for pilots, various findings were observed. Two studies found positive effects of the CDTI application. DiMeo (2001) tested out four different CR conditions including current operation, pilots with CDTI and two share-separation schemes. ATCOs indicated that current operation with the addition of CDTI could lead to more safe operations and lower workload. In Canton, et al.'s (2005) study, the CDTI was integrated with a conflict detection and resolution (CD&R) tool. An experimental trial was conducted and the overall result indicated that the CD&R-capable CDTI was successfully proven, although the concept of conflict resolution was not matured yet.

However, Dao, et al. (2011) examined three different CR concepts on pilots' situation awareness (SA) (pilot primary, controller primary, and auto-resolver primary) and no significant difference was found across the concepts. Subsequently, another study (Vu et al., 2012) examined the effects of the concepts in Dao et al. (2011) on pilots' performance and workload. Providing an automated separation assurance system to flight decks did not benefit pilot workload. Performance data also indicated that the efficiency of aircraft control, including time to top of descent and spacing lag, was worst when the pilot was equipped with an auto-resolver system.

Taken together, it can be inferred that the results from several studies examining the concept of CRA varied. Although there are clear empirical evidences on the benefits of the CRA as described in the majority of the studies, however, there are no consistent findings on whether ATCOs would appropriately utilize the CRA. Ultimately, no findings were found about how the CRA would affect the ATCOs' workloads, performances and SA, especially, when it errs, and how to mitigate the possible costs that may appear due to

the CRA imperfection. These research gaps and the series of experiments to address these gaps are presented further in the subsequent chapters of the dissertation.

2.4. Limitations in Studies on Conflict Resolution in ATC

Unlike in conflict detection, the effects of some HAI factors on ATCOs' conflict resolution process have received little or even no attention. Particularly, automation reliability which has been highlighted as the most important factor in many command and control fields and should be investigated for safe ATC operations, has never been examined in conflict resolution context. Given the worldwide increase in air traffic density and the nearly maximum capacity of the current ATC facilities, the possibility of air traffic conflict has become higher. This calls for research to obtain the knowledge concerning the issue of automation reliability in conflict resolution process.

Essentially, compared to conflict detection, conflict resolution involves higher cognitive processing which includes aircraft monitoring, attention switching, coordination, and instruction issuing (Kallus, et al. 1999). Because of these more challenging tasks, greater consequences may be expected if the process fails. However, as reviewed in Section 2.2.3, there are only three existing studies (Metzger & Parasuraman, 2005; Wickens, et al., 2009; Rovira & Parasuraman, 2010) that investigated automation reliability effects. However, these studies only focused on the conflict detection process and not resolution process. In addition, one limitation with Metzger & Parasuraman (2005) is that they always placed reliable automation condition before unreliable conditions. This resulted in longer conflict detection time for reliable automation than

unreliable automation but this effect was confounded with the order effect. Rovira and Parasuraman (2010) studied the conflict detection imperfection in terms of misses and false alarms. However, in their study, participants did not have voice communication as a means to coordinate with pilots and to interact with the system; the participants rather completed all tasks using keyboard and mouse. These task changes could cause a disturbance in ATC normal working procedures. Wickens, et al. (2009) performed an analysis of historical conflict alerting cases and found that ATCOs responded to true alerts although there was an inclination of non-responses due to false alarms. However, the data provided were only from the heaviest workload period within each centre and no direct experiment was conducted thus no direct access to participants was possible.

There are a number of HITL studies that examined conflict resolution automation as reviewed in Section 2.3.2 (Cardosi, et al. 1992; Dwyer & Landry, 2009; Prevot, Homola & Mercer, 2008; Prevot, et al. 2009; Jha, et al. 2011; Borst, Westin, & Hilburn, 2012; 2013; Cabrall, et al. 2014, and Goritzlehner, 2014). However, none of these studies had investigated the effects of automation reliability. All of these studies, except the studies by Cardosi, et al. (1992) and Borst, Westin and Hilburn (2012; 2013), examined perfectly reliable conflict resolution automation and compared it with manual performance. Cardosi, et al. did not compare with manual performance but only measured the time required for conflict resolution with perfect automation. Borst, Westin, and Hilburn (2012; 2013) investigated strategy conformance between ATCOs and automation without comparison with manual condition as the baseline.

The unreliability issue may also trigger different ATCOs responses depending upon the time available for them to take actions. The conflict risk due to unreliability and

its impact on ATCOs behaviors are crucial issues in need of further examination (Xu, 2003) and only Metzger and Parasuraman (2005) who had examined this issue in conflict detection. However, the time variation was a bit too long and did not give significant effects on the ATC operations.

Besides, the interaction effects between automation reliability and traffic density have not been studied. In fact, high traffic density has always been associated with poorer performance (Metzger & Parasuraman, 2001; 2005; 2006; Nunes, 2003) and high workload (Galster, et al. 2001; Willems, 2002). Traffic density is a factor that influences the use of automation (Westin et al., 2012) such that higher traffic would lead to higher dependence (Prevot, et al. 2008; Westin, et al. 2013). The only study that has investigated the interaction between automation condition and traffic density was conducted by Prevot, et al. (2012). However, they only varied the conflict resolution automation to be automated (perfectly reliable) and manual conditions. It is not clear from the Prevot, et al.'s study how the traffic density would appear to affect ATCOs' performance given imperfect automation. It is logical that higher traffic density would lead to more dependence on automation as an operator's cognitive resources are limited, since other research with alerts suggests that dependence increases with secondary task loading (Wickens & Dixon, 2007). Given the higher dependence upon automation, however, there may be a more serious consequence if automation makes an error and particularly the first time automation fails, as dependence may have been very high. The fatal consequence, however, is impermissible for passengers' safety.

Furthermore, integration of the vertical situation display in ATC facilities has not been examined alone, let alone in its interaction with CRA automation reliability has

never been explored. Vertical information is critical for ATC conflict resolution, given the ATCO's strong preference to instruct vertical maneuvers (Eyferth, Niessen & Spaeth, 2003; Rantanen & Nunes, 2005; Loft, et al. 2007; Rantanen & Wickens, 2012; Mercer, et al. 2014), and while the benefit of the vertical situation display in the CDTI for pilots (DiMeo, et al. 2001; Canton, 2005; Dao, et al., 2011; Vu, et al., 2012) has been demonstrated, in ATC, graphical depiction of vertical information has only been evaluated in the context of a 3D display (Wickens, Miller & Tham, 1996; Tham & Wickens, 1993), which has other costs in terms of ambiguity and integration with the lateral display. The VSD in plan (2D) format is examined here. Also, while display transparency has been found to offset automation imperfection costs in other domains (Bisantz & seong, 2001; Dzindolet, et al. 2003; Wang, et al. 2009; Nass, et al. 1996; Verberne, et al., 2012; Rovira, et al. 2014) by supporting situation awareness; this has not been examined in ATC. This is done in the current dissertation.

To address these limitations, three experiments will be conducted as the important parts of the continuously ongoing research effort on conflict resolution automation for future ATC operations and are elaborated in details in the following section. All three experiments will exploit information from an array of dependent variables in order to understand, not only the direct effects of the four independent variables on performance, but also the mediating effects of trust, dependence, situation awareness and workload.

Chapter 3

Research Framework

The overall focus of the dissertation involving three experiments (Figure 3.1) is to:

- Examine the ATC CRA (an imperfect automation decision aid)
- Assess its degree of support for controller performance, relative to unaided performance
- Assess the implications of imperfections in its reliability
- Examine how the above findings are influenced by time pressure and workload (manipulated via traffic load)
- Assess the assistance of a vertical situation display, to both improve overall performance, and mitigate any costs of imperfections in the CRA.

Figure 3. 1. Research Framework

Across the three experiments in which these issues were investigated, overall traffic conflict avoidance performance, workload, trust in and dependence on the aid, and situation awareness were measured. The reasons of selecting these human factors variable and the way to quantify the selected factors are described as follows:

3.1 Task performance measures

The existing ATC facilities are reaching the maximum capacity and will not be able to accommodate the projected increase in ATC (Airbus, 2013; ICAO, 2006; Sheridan, 2006). To deal with such increase, some programs (e.g. the U. S. Next Generation Air Transportation System (Cabrall, et al. 2014) and the Single European Sky ATM Research (SESAR, 2015)) that help implement new technologies and processes are under-development (Langan-Fox, et al. 2009). To achieve successful transformations of an overloaded system, evaluations of the transformation should be based on task performance with a new system and its comparison with the existing system as the baseline. By doing so, the safety, security, capacity, and efficiency of new system can be understood in a concrete way. Hence, measuring task performance, given a new system, is critical to ensure its benefits and that it does not cause higher cost for ATC operations.

The primary task for the participants was to resolve potential conflicts. To quantify the task performance measures, percentage of resolved conflicts, conflict resolution time and procedural errors were included. Percentage of resolved conflicts is operationally defined as the absence of loss of separation (LOS) relative to the number of designated conflicts following the CRA instructions. The separation standards for the experiment were 3 nautical miles laterally and 1000 feet vertically (Nolan, 1999). Conflict resolution

time is defined as the time interval between the CRA onset and the pilots' initial response. Errors in performing the ATC tasks (i.e. improper altitude clearance, inappropriate hand-offs, inappropriate runway assignments, and incorrect radio frequency changes) were counted. The ATCSimulator®2 is equipped with a tracking system of ATC tasks procedure. So long as participants did not meet the appropriate procedure, for instance, when participants gave a wrong radio frequency or wrong exit altitude, the system listed this particular violation in the reporting file.

In some cases specific performance was examined on the trial in which the imperfect aid failed to give appropriate resolution (i.e., the first failure) – to assess complacency, and on the trial following this failure, to assess the change in dependence that might result from encountering a failure.

3.2 Mental workload

Workload is arguably one of the most vital HF issues in the context of ATC (Rantanen & Nunes, 2005). Traffic density is associated with subjective (Galster, et al. 2001; Metzger & Parasuraman, 2001) and objective workload (Averty, et al. 2004). With the increase in air traffic density, ATCO's workload has become an even greater important issue. This is because the proximity of aircraft leading to loss of separation and the direction variations triggered by the traffic load. Other factor that contributes to ATCO's workload is also the introduction of revolutionary forms of automation that will increase workload considerably (Langan-Fox, et al. 2009).

Empirical research investigating the impact of increased automation on ATCO workload has provided diverse results. On one hand, prior studies (Prevot, et al. 2012;

Cabrall, et al. 2014) found that workload would be lower with the automation. However, on the other hand, information analysis and decision-making automation produced higher workload, due to the visual demands of displays. Despite some research suggesting reduction in ATCO workload with increased automation, there are other concerns about whether workload would actually decrease (Langan-Fox, et al. 2009). Sheridan (2002) explained that increased levels of automation require the human's supervision for the automation, which can impose considerably higher workload for even very simple tasks. Given these concerns and the new concept of CRA, the evaluation of the CRA on ATCOs' workload is essential.

The subjective workload measurement was obtained using the NASA-TLX (Hart & Staveland, 1988). The objective measures of mental workload were derived from ready response latency and percentage of timeouts in the SPAM (Loft et al., 2015). In this study, ready response latency was defined as the interval between a ready response onset and participants' response to the ready prompt. Percentage of timeouts was defined as non-responded questions.

3.3 Trust

Trust has been acknowledged as an increasingly important issue in the development and application of complex ATC system (Kelly, et al. 2003). In addition, trust is an essential construct so that proposed capacity and safety of an automated system can be fulfilled (Lee & Moray, 1992). However, there are limited researches on trust in ATC systems despite its importance. Trust in automation during the operations is often miss-calibrated where operators put too much trust or too little trust on the system. Both

phenomena have been observed in ATC and may lead to adverse effects on system and ATCOs' performance. Therefore, this dissertation assessed trust in the novel CRA system.

Subjective rating of trust in the CRA was obtained using the Likert-type rating scale (ranging from 0 to 7) developed by Jian, Bisantz, and Drury (2000). The questionnaire was administered after each testing condition.

3.4 Aid utilization rate

Dependence on the automation system is an important construct in ATC. ATCOs who rely much on the automated system can be over-complacent (Parasuraman, et al. 2000). Over-complacency could restrict ATCOs' understanding of the automation and increase the possibility of failure in monitoring the state of automation. Therefore, complacency on the automated system may lead to failure in detecting automation' error and it is more likely to occur in highly reliable system. In contrast, less reliance on the automated system could also impose a risk for maintaining ATCO's performance. ATCOs may monitor the system too frequently or simply do not want to utilize the system, thereby negating the intended benefits of the automation. The understanding on how ATCOs would respond to the proposed CRA is obtained by assessing their dependence on the proposed aid.

Aid utilization rate data that reflected the ATCOs' dependence upon the CRA were obtained through participants' acceptance of the advisories. During the simulation, two buttons (i.e. agree and reject buttons) appeared following the advisories. If participants opted to agree, the word 'agree' appeared on the pseudo-pilot screen and the pseudo-pilot would send a preset macro for resolving the respective conflict. If

participants pressed the reject button, they were asked to provide their own resolutions manually to the pseudo-pilot using voice transmission. Their responses to all advisories were recorded. The ratio of the number of agreement with the CRA to the total number of advisories was used to reflect participants' dependence upon the CRA.

3.5 Situation awareness

One of the major challenges for the introduction of the future ATC operational concepts is its impact on ATCOs' SA (Langan-Fox, et al. 2009). ATCOs' SA is referred to as ATCOs' "picture" that contains the perception, comprehension, and projection of an environment's elements (Endsley, 1995). Prior research has revealed the benefits of high level automation on SA (Endsley, et al. 1999) in some extent. However, research also found that high levels of automation could compromise the maintenance of system's awareness in a dynamic environment (Kaber, et al. 2004) such as ATC. The decline in SA could be associated with some factors such as increased monitoring demands, reduced vigilance, complacency and lack of trust (Endsley, 1997). Subsequently, the SA reduction under new systems may result in out-of-the-loop (OOTL) performance problems, particularly when the automated system fails (Wickens, et al. 2013) and lead to dangerous consequences. It is clear that future ATC concept must place an emphasis on the issues to maintain appropriate level of ATCOs' SA.

SA data were collected during the entire experiment by giving SA probes in two steps. First, a ready prompt was shown. This prompt informs participants that a probe question is ready to be presented (see Figure 3.2). The examples of probe questions, all 2 alternative forced choices, were adopted from the requirements of SA information for en-

route air traffic control (Endsley, 1994; Endsley, 2000) and provided in Appendix B. If participants felt capable of answering the probe question without affecting their ATC tasks, they could proceed to the question by pressing an “answer” button. Participants were able to see the questions right after pressing the button. The prompt appeared every 6 minutes (i.e., 9/condition), and appeared at random times relative to the conflict. The probe questions appeared right after clicking the answer button. The SA display timed out after one minute if participants did not respond to either the ready prompt or the probe questions.

Figure 3. 2 Example of a SA probe

SPAM ready response latency indicated the readiness of participants to answer SA questions and often correlates with subjective workload (Strybel, et al., 2008; Vu et al., 2012). Once participants indicated that they were ready, the accuracy and time taken to then answer the SPAM queries reflect SA (Durso & Dattel, 2004; Durso, et. al. 2006).

Hence, two SA measures were recorded (a) probe response latency defined as the interval between the display of a question and participants' answer, and (b) percentage of correct responses to SA probes.

The Experimental Studies

In Experiment 1, manual performance is compared to CRA-supported performance at two levels of CRA reliability: perfectly reliable (100%) and imperfect reliable (80%), and two levels of time pressure in resolving the conflict are also imposed.

In Experiment 2, the same general design as in experiment 1 is employed, but both the overall level of workload (traffic load) will be increased and that traffic load will be varied between conditions.

In Experiment 3, manual control is again compared to the same two levels of automation reliability (perfect and 80%) but the availability of a vertical situation display is also varied, designed to both improve overall performance, and to mitigate the possible effects of CRA imperfection as shown in Figure 3.3.

Figure 3. 3. Research model for experiment 3

In each of the following three chapters, the experimental details are presented. The experimental effects, revealed in testing the proposed hypotheses, are designed to help fill in the gaps of the literature; both pertaining specifically to the implementation of the CRA, and more generally to key issues in human-automation interaction.

Chapter 4

Study 1: Effects of Conflict Resolution Aid Reliability and Time Pressure on Human Performance Variables of Air Traffic Controllers

The current study aimed to evaluate the CRA from a human factors perspective by investigating the effects of CRA reliability and time pressure. It is important to realize that automated conflict resolution, like conflict detection, can also be unreliable, although in contrast to conflict detection, where misses and false alerts are easy to categorize as automation failures, the issue of what is the “wrong” advice (an automation conflict resolution failure) is harder to specify unambiguously. For example, is it “wrong” if the CRA instructs a speed maneuver, when the ATCO prefers to climb, even if both are equally effective? The ATCO might judge the CRA to be “wrong” in this instance, hence degrading its perceived reliability. No empirical data are available regarding the effects of automation reliability on ATCOs performance with automated conflict resolution aids. In fact, reliability and its effects on trust are arguably the most critical factor in any automated systems (Lee & See, 2004). In addition, the factor of time pressure imposed on ATCOs in an unreliable condition may affect ATCOs performance due to speed-accuracy trade-off (Kantowitz and Sorkin, 1983). Therefore, there is also a need to investigate how ATCOs utilize automation aids when they are exposed to different time pressure conditions. The independent variable was automation conditions which was manipulated by varying the CRA reliability levels with time pressure nested in the unreliable condition. The dependent variable was ATCOs’ performance, workload, SA, trust and dependence. This study tested seven hypotheses which are formed in Chapter 1.

4.1 Methods

4.1.1 Participants

Twelve college students (ten males and two females) aged 23 to 27 years (Mean = 25.08 years, SD = 1.38 years) participated in the present study. These participants were educated in Aerospace and Aeronautical Engineering from local institutions and averagely had two weeks ATC simulator experience. An interview was conducted before the study to make sure that all these participants were familiar with the procedures and terms used in the current ATC practice. Informed consent approved by NTU IRB was provided by each participant.

Prior to the formal experimental study, a power analysis in repeated-measure designs using the data from the pilot study as suggested by Maxwell & Delaney (2004) was performed. From the study, the effect size, d , was computed as follows:

$$d = \frac{\mu_{\max} - \mu_{\min}}{\sigma} = \frac{96.00 - 62.00}{19.99} = 1.70$$

The number of levels indicated the number of testing conditions of the studies. In the dissertation, the number of levels was 4. The minimum correlation found from the study was 0.32. With the d value of 1.70, referring to the power analysis table (shown in Appendix I), 12 participants were required to achieve 80% power with $\alpha = 0.05$.

4.1.2 Apparatus

A desktop version of ATC simulator, ATCSimulator®2, was used in the experiment. This simulator consisted of a radar display and electronic flight progress strips as shown in Figure 4.1. The radar showed a 60-mile radius sector. An ATC-Sector

Design Kit (ATC-SDK) was used to generate the traffic scenario in a generic airspace. The conflicts in each scenario were generated by manipulating the flight time, speed, and locations of aircrafts.

Figure 4. 1 Primary radar display

The experiment setup consisted of one air traffic controller (ATCO) position and one pseudo-pilot position that were connected to the main processor. In each position, there were two screens, including a 24-inch monitor that was used to display the primary radar and a 17-inch monitor that displayed the CRA (Figure 4.2). There are several reasons why the CRA was put on a separate screen: (i) it can minimize “display clutter” (Wickens, Kroft, & Yeh, 2000) on a primary radar display, (ii) the location of a monitor display (at or beside a central location) does not affect monitoring pattern and performance (Singh, Molloy, & Parasuraman, 1997), (iii) multiple displays are common

in the current ATC practice and research, for example the URET screen (Wickens, 1998) , the Multi-Attribute Task Battery, and the ATC Simulation with Flight Strips (Galster et al., 2001; Metzger & Parasuraman, 2005; Parasuraman et al., 1993).

The CRA provided advisories to participants. An example of these advisories was shown in Figure 4.2. This example advised the pilot of aircraft SWA356 to descend and maintain at ten thousand feet while the pilot of aircraft N130LM were asked to maintain its current course. The list of possible abbreviations in the advisory is provided in Appendix A. The CRA was based on the principle proposed by Erzberger (2006) for Resolution Aircraft and Maneuver Selector (RAMS), since regular patterns of conflict resolution maneuver appeared to be empirically verified (Rantanen & Wickens, 2012). The CRA adopted the altitude-first resolver principle where a vertical maneuver would be preferred over lateral and speed ones. The timing data of each conflict was inputted into the CRA, and the CRA appeared two minutes prior to any detected conflicts (i.e., loss of separation). The conflict resolution advisories were translated into macros containing resolution maneuvers in the radar simulation. Participants used a headset with a built-in microphone to communicate with a pseudo-pilot. A mouse was provided for the interaction between participants and the onscreen CRA control.

Figure 4. 2 CRA display

The CRA design was chosen based on two reasons; first, common practice in the current ATC display. The display uses common ATC terms in the current practice. The display indicates a maneuver with an abbreviation of the first character of a particular maneuver in the common ATC practice (Cardosi, 1992). In addition, displays used in the current ATC workplace, such as flight strips and User Request Evaluation Tool (URET) information are provided in the “listing” style (Endlsey, 1999; Wickens, 1998). Therefore, this increased ATCOs’ familiarity with the proposed CRA.

Second, display clutter on the primary radar should be minimized. The advisory sign is not placed directly on the radar display since it can lead to display clutter. It would lead to a great cost since ATCOs need to control many aircraft at the same time. Any cost on the ATCOs’ attention due to display clutter must be diminished. Signage indicated by arrows or other symbols representing a maneuver were not applied because such cues

require more cognitive resources to interpret (Wickens, et al. 2013). Hence, it takes a longer time to process, which cannot be compromised in any ATC practice.

4.1.3 Experimental design and procedure

A within-subjects design was adopted. In order to assess the role of CRA reliability and time pressure in ATC, four testing conditions were examined in the experiment:

Condition R: reliable CRA

Condition UAH: unreliable and high time pressure

Condition UAL: unreliable and low time pressure

Condition M: manual (no-CRA).

In order to remove any carry over effects, a balanced Latin square was used to order these testing conditions across participants. The order of the conditions is shown in Table 4.1.

Table 4. 1 The order of experiment conditions

Participant #	Testing Conditions			
1	R	UAH	UAL	M
2	UAH	M	R	UAL
3	UAL	R	M	UAH
4	M	UAL	UAH	R
5	R	UAH	UAL	M
6	UAH	M	R	UAL
7	UAL	R	M	UAH
8	M	UAL	UAH	R
9	R	UAH	UAL	M
10	UAH	M	R	UAL
11	UAL	R	M	UAH
12	M	UAL	UAH	R

Before the formal experiment, all participants were provided with a training session. Participants completed two 30-minutes radar training (i.e. without and with the CRA). During the practice session, participants were told that some complex factors such as flow constraints, area boundary, and iteration process could be potential triggers for the CRA aid to err. All participants were able to successfully give clearance for all departing and arriving aircraft at the end of the training session.

In the formal experiment, there were four one-hour ATC scenarios corresponding to the four testing conditions, respectively. Each scenario involved five potential conflicts. In each scenario, participants were instructed to communicate with the pseudo-pilots to issue appropriate altitudes, to accept all aircraft that entered their sector, to hand-off aircraft that left their sector and to issue the correct radio frequency change. The four scenarios were completed in four different experimental sessions with an at least one-day interval between each. In order to simulate future air traffic situations, sector density for each scenario was 30 aircraft. The experimenter acted as the pseudo-pilot during the experiment. After each CRA session, trust was assessed.

In the three CRA conditions, there was a conflict resolution advisory for each conflict and participants were free to either accept or reject the advisory by clicking a respective button. If participants agreed with the advisory, the word “agree” would appear in the pseudo-pilot position and the pseudo-pilot would directly execute the resolution maneuvers by giving the preset macros on the radar simulator. Otherwise, participants had to give their own resolution maneuvering advisory to the pseudo-pilot.

In the reliable condition, the CRA was able to provide correct advisories to all the potential conflicts. In the unreliable and high time pressure condition (UAH), the

maneuvering advisory helped resolve a primary conflict but led to a conflict with another aircraft (secondary conflict) in 100 seconds. The selected time interval between the CRA resolution of the primary conflict and the occurrence of the secondary conflict in the high time pressure condition was determined as the sum of the activation time of an onboard traffic alert and collision avoidance system (TCAS) (~ 45 seconds) (Thomas, Wickens, & Rantanen, 2003), necessary time for an ATCo to give a maneuvering instruction until a pilot response (~ 42 seconds) (Cardosi, 1991; 1992), and some time allowance (~ 10 seconds). It should be noted that once a TCAS is activated, ATCOs cannot interfere since the TCAS will give expedited maneuvers to avoid the conflict. Thus, setting a secondary conflict in 100 seconds after resolution of a primary conflict is based on the minimum time available for ATCOs to resolve a conflict in the current system which will surely impose time pressure onto them. In the unreliable and low time pressure condition (UAL), the maneuvering advisory elicited a secondary conflict in four minutes.

It should also be noted that in the unreliable conditions, if participants reject the advisory given by the CRA and instruct their own appropriate resolution maneuver, the secondary conflict could be avoided. However, if participants accepted the advisory in the conflict where a failure had been set, the advisory would resolve the primary conflict but trigger a secondary conflict. This form of automation failure was chosen, rather than a “failure” in the resolution of the primary contact, because of the ambiguity in precisely defining the latter.

An example of such a failure is given in Figure 4.3, in which the CRA correctly advised the primary conflict resolution between SNP567 and N80866. But such an advisory led to a secondary conflict between N80866 and FRL789 in 100 seconds. In the

manual condition, participants were asked to resolve the potential conflicts by providing their own resolution maneuvering instructions.

Figure 4. 3 Example of a secondary conflict situation

4.1.4 Dependent Measures

The dependent measures were performance, workload, trust, aid utilization rate, and SA. The definition of the measures and the quantification are listed in Chapter 3.

4.1.5 Statistical Analysis

The conflict resolution time, trust ratings, and aid utilization rate were only collected in the three CRA conditions. Therefore, the data were analyzed using a repeated measure ANOVA with one three-level independent variable and two planned orthogonal contrasts: (a) reliable versus unreliable conditions, (b) unreliable conditions with high versus low time pressure to resolve a secondary conflict.

The remaining dependent variables (i.e. percentage of resolved conflicts, procedural errors, mental workload, and situation awareness measures) were collected in all the four scenarios. The data were analyzed using a repeated measure ANOVA with the CRA reliability as the four-level independent variable and three planned orthogonal contrasts: (a) three automated versus manual conditions, (b) reliable versus two unreliable conditions, (c) unreliable conditions with high versus low time pressure to resolve a secondary conflict for respective dependent measures as appropriate. A separate contrast between manual and unreliable conditions was also conducted. Level of significance (α) was set at 0.05 for all the analysis.

Prior to the analyses, the assumptions for repeated measures ANOVA were tested including normality of residuals and sphericity prior to the analyses, as suggested by Montgomery (2008). The tests revealed a violation on the assumption of normality

of residuals on few measures (i.e. conflict resolution performance and aid utilization rate). ANOVA is robust against violation in normality and the F-test is found to be remarkably insensitive to the normality violation (Ito, 1980). Especially when the group sample sizes are equal. However, a natural logarithmic transformation was performed on the dependent variables which violated the assumption, and it remedied the violation. The sphericity assumption of the repeated measures ANOVA was evaluated using Mauchly's sphericity test, and Greenhouse-Geisser adjustment was adopted in case of a violation.

In addition, a significance of difference in proportions test was also performed on the aid utilization rate following an automation failure in the first unreliable block (whether UAH or UAL) and following a correct automation in the reliable block. The conflict resolution time for the trial following the automation failure was also analyzed. These analyses were done to reveal any loss of utilization that might have occurred as a consequence, reflecting a loss of trust.

4.2. Results

4.2.1 Task performance measures

There was a significant effect of CRA reliability on the percentage of resolved conflict (Table 4.2) with the effect size (η_p^2) of 0.541. Figure 4.4 shows the comparison across the four conditions. The contrast analyses revealed that more conflicts were accurately resolved under the automated conditions ($M = 88.33\%$, $SE = 2.67\%$) than the manual condition ($M = 55\%$, $SE = 7.80\%$) ($F(1, 33) = 40.59$, $p\text{-value} < 0.001$, $r_c = 0.69$). Conflict resolution performance was also better under the reliable condition ($M = 98.30\%$,

SE = 1.70%) than the unreliable conditions (M = 83.40%, SE = 3.15%) ($F(1, 33) = 7.31$, p -value = 0.011, $r_c = 0.29$). Surprisingly, participants under the unreliable with high time pressure (UAH) condition (M = 90.00%, SE = 3.00%) could resolve conflicts more accurately compared to the unreliable with low time pressure condition (M = 76.70%, SE = 3.30%) ($F(1, 33) = 4.33$, p -value = 0.045, $r_c = 0.22$). A final contrast between the unreliable conditions and manual condition showed that more conflicts could be more accurately resolved under the unreliable conditions (M = 83.40%, SE = 3.15%) than the manual condition (M = 55%, SE = 7.80%) ($F(1, 22) = 18.98$, p -value < 0.001), verifying that ATCOs performance was better under even unreliable automated conditions than manually performing the tasks. No significant difference in CR accuracy was observed between the FF conflict (M = 83.33%, SE = 11.24%) and the manual performance (M = 58.33%, SE = 14.86%) ($z = 1.35$, p -value = 0.18, $r = 0.19$), suggesting that complacency effect was absent in this study.

Figure 4. 4 Conflict resolution performance (error bars indicate +/-1 SE)

A significant effect of CRA reliability on the conflict resolution time was observed (Table 4.2). The effect size for the F-test (η_p^2) was 0.274. Conflicts were resolved faster under the reliable condition ($M = 46.75$ s, $SE = 6.49$ s) than the unreliable conditions ($M = 70.58$ s, $SE = 10.47$ s) ($F(1, 22) = 7.75$, $p\text{-value} = 0.011$, $r_c = 0.51$). There was no effect of time pressure on conflict resolution time. Importantly, there was a significant delay in conflict resolution time after participants experienced automation failures ($M = 103.84$ s) compared to that in the reliable condition ($M = 46.75$ s) ($F(1, 11) = 24.01$, $p\text{-value} < 0.001$, $r_c = 0.83$). No significant effects of CRA reliability on the procedural errors were observed (Table 4.2).

Table 4. 2 Dependent measures under the various CRA conditions: Mean (SE)

Measures	CRA Conditions				$p\text{-value}$	F-value
	Reliable	Unreliable and High Time Pressure	Unreliable and Low Time Pressure	Manual		
Percentage of resolved conflict (%)	98.30 (1.70)	90.00 (3.00)	76.70 (3.30)	55.00 (7.80)	0.002*	12.94
Conflict resolution time (s)	46.75 (6.49)	74.25 (11.02)	66.92 (9.91)	-	0.030*	4.15
Procedural error (number)	2.50 (0.54)	2.08 (0.34)	2.33 (0.41)	3.17 (0.44)	0.288	1.31
Trust (ratings)	6.29 (0.14)	6.08 (0.17)	5.92 (0.20)	-	0.080**	2.88
Workload (NASA TLX)	74.00 (3.70)	71.77 (3.40)	71.50 (3.40)	74.06 (3.70)	0.395	1.02
Aid utilization rate (%)	93.33 (3.76)	85.00 (8.21)	88.33 (6.72)	-	0.983	0.02
Percentage of timeouts (%)	13.50 (6.00)	16.30 (4.50)	21.70 (6.30)	15.10 (3.80)	0.497	0.81
Ready response latency (s)	8.00 (0.98)	9.50 (1.82)	9.67 (1.78)	9.08 (1.59)	0.832	0.29
Probe response latency (s)	14.83 (1.59)	15.25 (1.35)	14.08 (1.46)	17.00 (1.45)	0.254	1.42
Percentage of correct response to SA probes (%)	79.80 (2.90)	84.80 (2.90)	85.30 (4.00)	69.30 (6.40)	0.037*	3.17

4.2.2 Subjective Ratings of Mental Workload and Trust

Participants rated their workload to be quite high ($M = 70.28$; $SE = 3.60$), compared to a maximum possible value of 100. However, no significant difference in mental workload was found between the four conditions, $\eta_p^2 = 0.09$ (Table 4.2).

The trust ratings ranged from 1 to 7 with an average of 6.02 ($SE = 0.17$) indicating a fairly high trust on the CRA. The effect of CRA reliability on trust ratings was only marginally significant with the effect size (η_p^2) of 0.224 (Table 4.2). The orthogonal contrast showed that participants rated their trust higher under the reliable condition ($M = 6.29$, $SE = 0.14$) than under unreliable conditions ($M = 6.00$, $SE = 0.19$) ($F(1, 22) = 4.71$, $p\text{-value} = 0.041$, $r_c = 0.41$). The contrast showed no significant effect of time pressure.

4.2.3 Aid Utilization Rate

There was no significant effect of CRA reliability on the aid utilization rate (Table 2). Participants tended to depend upon the aid to resolve most of the conflicts ($M = 88.89\%$, $SE = 6.23\%$) regardless of the levels of CRA reliability. Although it is lower, the dependence on the CRA following a CRA failure ($M = 83.33\%$) was not significantly different with that following a correct CRA ($M = 93.33\%$) ($z = 0.73$, $p\text{-value} = 0.465$, $r_c = 0.15$) in reliable condition.

4.2.4 Situation Awareness

A significant effect of CRA reliability on the percentage of correct responses to SA probe was found with the effect size (η_p^2) of 0.224 (Table 4.2). The contrast analysis

revealed that the response accuracy was higher under the automated conditions ($M = 83.30\%$, $SE = 3.27\%$) than the manual condition ($M = 69.30\%$, $SE = 6.40\%$) ($F(1, 33) = 8.43$, $p\text{-value} = 0.007$, $r_c = 0.45$). It was also found that unreliable automation ($M = 85.05\%$, $SE = 2.30\%$) led to higher response accuracy than the manual condition ($M = 69.30\%$, $SE = 6.40\%$) ($F(1, 22) = 7.67$, $p\text{-value} = 0.011$, $r_c = 0.50$) (Figure 4.5) The effect of CRA reliability on probe response latency was not significant (Table 4.2). There were no other significant results.

Figure 4. 5 Percentage of correct SA response (error bars indicate +/- 1 SE)

4.3 Discussion

The primary purpose of the experiment was to examine the effects of automation reliability and time pressure within the framework of performance, workload, situation awareness, automation trust and dependence. In the current study, conflict resolution automation was evaluated.

The first and the second hypotheses were partially upheld. The results showed that the CRA clearly helped ATCOs in improving resolution rates by 33% as compared to manually performing the task. In addition, perfectly reliable CRA also improved performance above unreliable CRA from 83% to 98%. However, participants' workload was not reduced by the automation. Workload was also not different across the automation conditions. This could be seen from the non-significant results on the workload rating (Table 4.2). These findings could be explained in the way that the CRA supported ATCOs in performing their tasks, but they still needed to attend to other tasks that were not supported by the CRA, making their workload remained unchanged.

For Hypothesis 3, the result in Table 4.2 showed that even imperfectly reliable automation substantially increased resolution performance to more than 80%, despite the non-different workload levels, demonstrating that this hypothesis was also partially supported. It is noteworthy that this performance assistance is consistent with the fact that the reliability is above the Wickens & Dixon (2007) threshold. In addition, there was no cost on the workload level due to the unreliability, suggesting that ATCOs remain in the loop when the off-nominal event occurred.

The result of significantly higher SA in automated conditions than in manual condition (Table 4.2) demonstrated that Hypothesis 4a was upheld. This increase in SA

accuracy with the CRA is actually consistent with the workload finding. Participants apparently deployed any freed resources, availed by automation, to pay better attention to the “raw data” on the radar display, hence increasing SA.

Regarding H4b, it was observed that while imperfection of the CRA did degrade performance somewhat from the fully reliable condition, in both speed and accuracy, it also improved SA in a way that partially offset the costs of that imperfection. This showed that H4b was also upheld. Presumably, becoming slightly skeptical or mistrustful of the automation on these imperfect blocks based on trials following the first failure, participants paid closer attention to the raw data depicting possible secondary conflicts hence increasing their accuracy of response to SA probes (Figure 4.5); but they could still greatly benefit from the 100% perfection of automation in recommending resolution from the primary conflicts.

Hypothesis 5 was partially upheld. As hypothesized, trust in the CRA declined with reliability. Trust was lower in unreliable than in reliable condition. However, dependence, as assessed by utilization rate was high (89%) and only slightly decline with reliability. Thus, participants noted the automation failures, adjusted their SA and trust accordingly, but still highly depended or relied upon automation, presumably because the high workload task environment allowed them little other alternative (Wickens, et al., 2013).

In the above analysis, performance, reliability and use/dependence were averaged across an entire block. Of course this average masks conflict-by-conflict variations in these measures – the dynamics of trust and dependence – the most important of which is the response following the initial failure experienced by each participant. Because of the

counter balancing design, this initial failure occurred at a different trial blocks for different participants. Upon realizing that a prior perfect CRA had now failed, did this lead to a temporary drop in trust and dependence? Because trust was only measured at the end of a block, it could not be assessed. However, trust degradation could be characterized by more attention paid to the raw data which induced longer response time (Wickens, Conejo, & Gempier, 1999).

Following automation failures, in the present study, participants trust was found to degrade, indicated by the significantly longer conflict resolution time to the first conflict trial after the failure trials. Also, the likelihood of ignoring the automation following an automation failure as compared to that following a correct automation was examined. The results revealed that participants showed less dependence after automation failure encounters (83.33%) than after correct trials following a correct automation (90%) although it was not significant. These findings showed that participants' trust was fairly calibrated although they still depended upon the automation. The phenomenon might occur due to high workload so that they opted to rely on the CRA to preserve their cognitive resources for other tasks (Wickens & Dixon, 2007). Also, participants might depend on the CRA upon the initial contact (i.e. the first experience of using such aid), because they were not really familiar with the system (Merritt & Ilgen, 2008).

Hypothesis 6 was not supported. The results showed significant difference between both high and low time pressure condition on performance but not on workload (Table 4.2). However, it was surprising that participants with more time to resolve the secondary conflict did not perform better than those having less time to act in response. This may be due to the nature of ATC practice which involves tactical actions (Hoffman,

Mukherjee, & Vossen, 2010). This means that a secondary conflict that was triggered within four minutes may reflect a really unexpected event because participants may have switched their attention away from that particular possibility. In contrast, the high time pressure condition represents a condition in current ATC practice where a traffic alert typically appears (Thomas, Wickens, & Rantanen, 2003). Thus, participants may perform better since they are more familiar with it.

Furthermore, time pressure did not affect participants' mental workload. "Workload insensitivity" (Parasuraman & Hancock, 2001) might have occurred in the conflict resolution task, where ATCOs could adapt to the unreliability making the workload remain unchanged, regardless of the time pressure.

Time pressure did not affect situation awareness in the present study indicating that H7 was not supported. These findings conflict with previous studies (Endsley & Garland, 2000; Gibson, Orasanu, Villeda, & Nygre, 1997) revealing that time pressure was perceived as one of the critical components in the situation assessment phase and affected 45% of pilots' situation awareness errors. However, some researchers also reported that time pressure and awareness levels were not related to each other (Rodgers, Mogford, & Strauch, 2000). The finding regarding the effects of time pressure on situation awareness might be because both 100 seconds and 4 min are within the tactical time range (i.e. less than 5 minutes) (Thomas, Wickens, & Rantanen 2003), in which participants might expect that there would be a secondary conflict because of CRA unreliability. Similar finding was found in Metzger and Parasuraman (2005) that reported no variation in conflict detection between 2 minutes and 4 minutes time ranges before a secondary conflict occurs.

Since conflict resolution requires more processing resources on the prediction of intervention's consequences (Eyferth, Niessen & Spaeth, 2003), the above results showed that the CRA could help highlight the consequences accordingly and minimize the demand on processing resources. In the present study, the examination of incorrect resolution with traffic aircraft might turn the totally unexpected event of primary conflict into the moderately unexpected event of secondary conflict (Wickens, Hooey, Gore, Sebok, & Koenicke, 2009). This means that a primary conflict which may be more difficult to detect by participants can successfully be resolved by the CRA, leaving participants with a "somewhat unexpected" secondary conflict. Given this, participants would be able to better resolve secondary conflicts because they have an initial knowledge on the predetermined possible conflicts. This reveals the importance of initial conflict resolution reliability. Based on these findings, it might be argued that automation for conflict resolution will be useful to be applied in future ATC as long as it can be reliable for the first conflict resolution iteration.

Chapter 5

Study 2: Interaction of Conflict Resolution Aid Reliability and Traffic Density in Future Air Traffic Control

In the current research, the CRA aid that integrates the conflict detection and resolution system was used as a tool for investigating the above issues in human-automation interaction (HAI). This choice allows us to examine the imperfections in the CRA system, in a manner that has rarely been done within the very small body of research that has examined controller performance with the CRA (Cabrall et al., 2014; Prevot, Homola, & Mercer, 2008; Prevot et al., 2012; Trapsilawati et al., 2015). Only Study 1 examined this issue of imperfect CRA and it was indeed found that the imperfect CRA aid assisted ATCOs relative to manual performance. However in Study 1, task load was not varied to examine if this factor amplified CRA aid benefits and imperfection costs. Thus the current study extends the general paradigm employed in Study 1 to examine the issues of traffic density. The independent variables were automation conditions and traffic density. Similar to Study 1, the dependent variables were ATCOs' performance, workload, SA, trust, and dependence. Study 2 aimed to test eight hypotheses as listed in Chapter 1.

5.1 Methods

5.1.1 Participants

Twenty-four participants aged 19 to 34 years ($M = 24$ years, $SD = 3.33$ years) were recruited. The participants in this study were different from those involved in the previous

study (Chapter 4). The participants consisted of 22 students and 2 professional ATCOs in Singapore. The students majored in Aerospace and Aeronautical Engineering from local institutions and had at least two weeks ATC simulator experience during their course. An interview was conducted before the study to make sure that all these participants were familiar with the procedures and terms used in the current ATC practice.

5.1.2 Apparatus

Apparatus used in Study 2 was identical to that used in Study 1.

5.1.3 Design

A mixed-factorial design was adopted. The first factor, automation level was a within-subjects factor with four levels: reliable, unreliable and high time pressure (UAH), unreliable and low time pressure (UAL), and manual (Table 5.1). The four automation levels were presented across participants in a sequence specified by a balanced-Latin square method. The second factor, traffic density was a between-subjects factor with two levels: medium vs. high. Sector density for medium and high traffic density was 60 and 90 aircraft, respectively. The participants were randomly assigned to these two traffic density conditions with each condition having twelve participants. The dependent measures were similar as Study 1 and listed in Chapter 3.

The conflict scenarios were similar across the four testing conditions. However, the aircraft call-signs, waypoints and occurrence times were changed. The traffic patterns were also rotated. This produced generally similar scenarios across the testing conditions to ensure that other factors such as conflict scenario (Thomas & Rantanen, 2006) would

not influence the effects of manipulation in different testing conditions. Each scenario was approximately 60 minutes in length.

Table 5. 1 Experimental Design

Condition	Reliable	UAH	UAL	Manual
High Traffic	X	X	X	X
Medium Traffic	X	X	X	X

5.1.4 Tasks

Each participant performed ATC tasks with the goals of maintaining separation and controlling the traffic flow. Participants used voice transmission to communicate with a pseudo-pilot. Participants had to handle all arriving and departing aircraft within their controlled area. For arriving aircraft, participants needed to: (1) accept incoming aircraft to their area, (2) clear the aircraft altitude to the landing altitude, (3) assign an appropriate approach clearance and runway, either Instrument Landing System (ILS) or Visual Approach, and (4) hand-over the aircraft to the tower controller. For departing aircraft, participants were responsible for (1) climbing the aircraft to the exit altitude, and (2) handing-over the aircraft to sector facilities.

Upon receiving voice instructions from the controllers, the pseudo-pilot inserted respective commands to the ATC simulator. The ATC simulator used in the study provides synthetic voices that met standard phraseology used in the real ATCO-pilot communication procedures when the pseudo-pilot inserted any instructed commands through keystrokes.

In the manual condition, participants were required to perform all the above tasks manually, without the CRA. In the automated conditions, the participants had available recommendations of the CRA. Six conflicts, including five pre-set primary conflicts and one secondary conflict, were imposed in the unreliable conditions. The detection of primary conflicts was 100% accurate and led to the automated resolution recommendation at 2 minutes prior to the LOS. Participants interacted with the onscreen CRA control using a computer mouse connected to the system. Participants could either accept or reject the resolution advisories provided by the CRA by clicking accept or reject button, respectively.

If the ATCOs accepted the advice, the resolution advice would be automatically sent to the pseudo-pilot's screen. The pseudo-pilot would directly apply the resolution by giving the preset macro commands to the simulator. The conflict resolution would always be effective to successfully resolve the primary conflict. If the ATCOs rejected the resolution advice, the CRA would stop processing the respective aircraft's data, and would not be triggered again if the ATCOs implemented an ineffective resolution. The CRA would not have enough time to be triggered (i.e. less than 2 min) and generate a resolution for any ineffective resolution made by the ATCOs. This situation then was counted as an unresolved conflict for the conflict resolution performance.

Participants were also requested to respond to the SA probes during the experiment. Each probe appeared every 6 minutes; hence there were 9 probe questions in a one-hour scenario. Seven probes appeared before the first CRA failure (i.e. the fourth conflict), 1 probe appeared during the CRA failure (in between primary conflict and secondary conflict which was within 4 minutes and 100 seconds under UAH and UAL,

respectively), and 1 probe appeared after the post failure trial. The ready and question probes would disappear after 1 minute of no response.

5.1.5 Automation Implementation

Participants were provided with a resolution advisory for each potential conflict in the three automation conditions. In the reliable condition, all advisories provided by the CRA were correct (100% of aid reliability). In the unreliable conditions, the aid reliability was 80% (1 failure out of 5 conflict trials). The advice for the fourth conflict helped resolve the primary conflict but led to a secondary conflict with a traffic aircraft and no advice was provided to avoid the secondary conflict. The failure was placed in this fourth conflict trial because it would allow participants to develop their trust in the CRA over the prior three trials, and for the primary conflict of the fourth trial before a failure occurred.

The decision to impose unreliable automation by the CRA missing a secondary conflict allowed for the implementation of automation error in automation advice while experimentally better controlling for the difference in maneuver preference between ATCOs. Furthermore, models for the development of on-ground CRA for ATC facilities are complex and require a large number of rules to completely cover all possible encounter situations and it may fail in certain situations. (Kuchar & Yang, 2000; Rantanen & Wickens, 2012). Hence, this study evaluated the failure situation that may likely happen in future ATC environment.

In the UAH condition, if the ATCO accepted the CRA, then it would successfully resolve the first conflict as noted above, but this resolution would trigger a secondary

conflict between one of the conflicting aircraft in the primary conflict and a traffic aircraft **100 seconds** after implementing the resolution (Figure 5.1). The traffic aircraft was intentionally preset in the scenario. For the secondary conflict, the CRA would not be triggered again, since this was the meaning of CRA imperfection in the study. The radar system generated an alert for the ATCO if a secondary conflict (loss of separation) occurred by a beeping sound as the evidence to the ATCO of the error.

Figure 5. 1 UAH condition

In the UAL condition, the secondary conflict would again be triggered between one of the aircraft in primary conflict and a traffic aircraft but now, this would be **4 minutes** after implementing the resolution maneuver (Figure 5.2).

Figure 5. 2 UAL condition

5.1.6 Statistical analysis

A 4(automation level) x 2(traffic density) mixed-design ANOVA was performed to analyze the percentage of resolved conflicts, mental workload, and SA. The trust ratings, CR time, and dependence were analyzed using a 3x2 mixed-design ANOVA since the data were only collected in the three CRA conditions. The collected data for all measures were tested for mixed-design ANOVA assumptions including normality of residuals, homogeneous variance of residuals, and sphericity prior to the analyses (Montgomery, 2008). The tests revealed a violation on the assumption of normality of residuals on few measures (i.e. conflict resolution performance and aid utilization rate). ANOVA is robust against violation in normality and the F-test is found to be remarkably insensitive to the normality violation (Ito, 1980). Especially when the group sample sizes are equal. However, a natural logarithmic transformation was performed on the dependent variables which violated the assumption, and the violation was treated after the transformation. The sphericity assumption of the repeated measures ANOVA was evaluated using Mauchly's sphericity test, and Greenhouse-Geisser adjustment was adopted in case of a violation.

In addition to the omnibus analyses, a series of planned orthogonal contrasts was performed to closely examine the effects of CRA and time pressure. First, the conflicts 1 to 5 were analyzed by pooling the data of all the trials to examine the effects of automation versus manual condition with the medium and high traffic densities. Second, to investigate the effects of CRA reliability, a contrast analysis (i.e. reliable vs unreliable) was performed exclusively on the fourth trial in the automation conditions for the conflict resolution time, dependence, ready response latency (i.e. workload indicator), and probe

response latency and accuracy (i.e. SA indicators). This trial was chosen because it was only here, in the two unreliable blocks, those participants would be aware that the automation was imperfect, having experienced the automation failure on the previous conflict. Hence, on this conflict, the direct effects of unreliability on dependence could be measured. Finally, a targeted analysis on the secondary conflict (occurring in trial 4) was performed to investigate the effects of time pressure on the resolution performance.

The statistical analyses around the FF trials were more complex. A series of analyses were performed to examine the FF effects. First, to examine whether FF hit rate was lower than that with unaided automation, a contrast analysis was performed between the FF proportions with the mean for the manual trials. Second, a test of proportions was performed to examine the FF effect, such that the FF effect is worse than subsequent failures. Third, a targeted analysis was performed exclusively because of the counterbalancing scheme (Table 5.2) in the Study 2, the participants could be subdivided into two groups depending on when they encountered the FF. Group E (for “early”) were given the unreliable automation (either UAH or UAL) in block 1, and hence received 3 reliable automation trials before the FF (which always occurred in trial 4 within the block) (Lines 2 and 3 in Table 5.2). Group L (for “late”) received this first unreliable trial in block 2, and hence had 8 perfect automation trials before the FF (Lines 1 and 4 in Table 5.2). By doing so, three aspects could be identified: (1) those observed on the FF relative to manual performance, which might reflect a complacency effect. (2) Those FF effects that differed between the FF and the subsequent failures as well as between groups E and L that might reflect the differential build-up of complacency (Molloy & Parasuraman,

1996), and (3) those that differed between pre-FF and post-FF trials that might characterize the recalibration of trust and dependence (Lee & Moray, 1992).

Table 5. 2 Counterbalancing of the Experiment Conditions

1	R (5)	UAH (3 1 1)	M (0)	UAL (3 1 1)
2	UAH (3 1 1)	UAL (3 1 1)	R (5)	M (0)
3	UAL (3 1 1)	M (0)	UAH (3 1 1)	R (5)
4	M (0)	R (5)	UAL (3 1 1)	UAH (3 1 1)

5.2 Results

5.2.1 Task performance measures

5.2.1.1 Percentage of resolved conflict

The main effect of automation level on the percentage of resolved conflicts was significant, $F(3, 66)=47.06$, $p<0.01$, $\eta_p^2=0.681$. All automation groups ($M=87.50\%$, $SE=4.24\%$) improved resolution performance above the manual condition where participants were not equipped with the CRA ($M=50.00\%$, $SE=5.34\%$), $F(1, 66)=103.98$, $p<0.01$ (Figure 5.3). More conflicts were successfully resolved under the medium traffic density ($M=87.08\%$, $SE=4.06\%$) than high traffic density ($M=69.17\%$, $SE=4.97\%$) conditions, $F(1, 22)=19.37$, $p<0.01$, $\eta_p^2=0.468$. A significant interaction effect between automation level and traffic density was observed, $F(3, 66)=8.61$, $p<0.01$, $\eta_p^2=0.281$, which confirmed the amplification of the CRA benefits under high traffic. Thus both H1 and H2 were strongly confirmed.

For the fourth trial, there was no performance difference between reliable ($M=95.93\%$, $SE=4.17\%$) and unreliable ($M=89.58\%$, $SE=4.46\%$) conditions, $F(1,44)=0.85$, $p=0.36$, $r_c=0.14$, indicating the benefit of even the CRA that had proven to be unreliable in the previous conflict.

For the targeted analysis of the secondary conflict, no significant difference between UAH ($M=54.17\%$, $SE=10.39$) and UAL ($M=70.83\%$, $SE=9.48\%$) was found, $F(1,22)=1.09$, $p=0.31$, $r_c=0.21$, although it is noted that both of these values are considerably lower than the average over all CRA-supported trials (approximately 85%), reflecting the costs of an automation failure.

Figure 5. 3 Conflict resolution performance across traffic densities (error bars indicate ± 1 SE)

5.2.1.2 Conflict resolution time

Conflicts were resolved significantly faster under the medium ($M=21.33s$, $SE=3.12$) than under high traffic density ($M=38.33s$, $SE=4.95$) $F(1, 22)=6.91$, $p=0.015$, $\eta_p^2=0.239$. There were no other significant results on resolution time.

5.2.2 Mental workload vs Subjective ratings of mental workload

Ready response latency in the SPAM indexed objective workload. There was a significant difference for ready response latency to suggest that automation ($M=9.25s$, $SE=1.68s$) did reduce workload than manually performing the tasks ($M=13.38s$, $SE=2.44s$) by shortening this latency, $F(1, 66)=7.50$, $p=0.01$, $r_c=0.33$. For the subjective workload measure, there were no significant results.

For the fourth trial, higher workload as indicated by longer ready latency was found under UAL ($M=11.48s$, $SE=2.63s$) than UAH ($M=5.95s$, $SE=0.84s$) condition, $F(1, 44)=5.25$, $p=0.03$, $r_c=0.32$.

5.2.3 Situation awareness

No differences between automated and manual conditions were found on probe accuracy ($M=70.21\%$, $SE=4.57\%$ vs $M=66.80\%$, $SE=5.07\%$), $F(1, 66)=1.01$, $p=0.32$, $r_c=0.12$ as well as on probe latency ($M=14.42s$, $SE=0.73s$ vs $M=15.08s$, $SE=1.03s$), $F(1, 66)=0.80$, $p=0.37$, $r_c=0.11$. There were no significant effects of traffic density on probe accuracy ($M_{High}=67.39\%$, $SE=7.85\%$ vs $M_{Med}=71.69\%$, $SE=5.25\%$), $F(1, 22)=0.42$, $p=0.52$, $\eta_p^2=0.019$ as well as on probe latency ($M_{High}=14.23s$, $SE=0.97s$ vs $M_{Med}=14.94s$, $SE=1.29s$), $F(1, 22)=0.29$, $p=0.59$, $\eta_p^2=0.014$.

For the fourth trial (that is, the conflict probe during the secondary conflict), there were no differences between reliable and unreliable conditions in probe accuracy to this 8th probe ($M=88.89\%$, $SE=5.68\%$ vs $M=87.09\%$, $SE=5.86\%$), $F(1, 44)=0.05$, $p=0.82$, $r_c=0.03$ as well as in probe response latency ($M=15.43s$, $SE=0.81s$ vs $M=18.52s$, $SE=1.84$), $F(1, 44)=2.17$, $p=0.15$, $r_c=0.20$.

5.2.4 Subjective ratings of trust

Trust was higher after the blocks in which participants were provided with reliable ($M=5.54$, $SE=0.21$) than unreliable CRA ($M=5.23$, $SE=0.23$), $F(1, 44)=3.99$, $p=0.05$, $r_c=0.28$. There were no other significant results.

5.2.5 Aid utilization rate

No difference in dependence between reliable ($M=79.17\%$, $SE=8.47\%$) and unreliable ($M=75.55\%$, $SE=7.37\%$) conditions was found, $F(1, 44)=0.17$, $p=0.68$, $r_c=0.06$.

5.2.6 First failure effect

In the FF trials, the conflict resolution performance was still better than that in the fully manual block (unaided performance) (72.22% vs 51.67% , $F(1, 35) = 16.14$, $p < 0.01$, $r_c = 0.56$), suggesting the absence of complacency effect and verifying the benefits of the CRA. Next, no significant difference in performance was observed between the FF

and the subsequent failures (69.44% vs 75.00%, $z = 0.53$, $p = 0.60$, $r = 0.07$), showing that the true FF effect was absent. In addition, those who encountered the FF “early” did however resolve the FF conflicts marginally significantly more rapidly (20 sec) than the “later” group (33 sec) ($F(1, 22) = 3.21$, $p = 0.09$, $\eta_p^2 = 0.127$).

In medium traffic density, the dependence rate on the CRA in the pre-FF trials was 91.67%. In the trials following the FF, automation dependence reduced to 83.33%. In higher traffic density conditions, the dependence rate in the pre-FF trials vs the post-FF trials was 83.33% and 75.00%, respectively. This supports the assumption that participants’ dependence was fairly calibrated. However, proportion tests of the pre-FF vs the post-FF in the medium density conditions ($z = 0.62$, $p = 0.54$, $r = 0.09$) and in the higher density conditions ($z = 0.50$, $p = 0.62$, $r = 0.07$) revealed no difference in dependence rates across the phases under both density conditions.

There was a significant effects of failure phase on workload $F(2, 22) = 6.08$, $p < 0.01$, $\eta_p^2 = 0.217$. There was a significant workload increase as indicated by ready latency, during post-failure ($M = 18.61$) compared to during pre-failure ($M = 9.83$), $t(23) = 3.61$, $p < 0.01$, showing the recalibration effects where participants tended to monitor the traffic closely after experiencing the automation failure, thus increased their workload.

The loss of SA during the FF was manifest in the SA probe response latency ($F(2, 22) = 16.76$, $p < 0.01$, $\eta_p^2 = 0.432$). The longest response latency was observed on the occasion of the FF ($M = 23$ seconds) compared to the pre- ($M = 14.26$ seconds) and the post-failure ($M = 12.8$ seconds) phases (Table 5.3). The interaction between groups (Early vs Late) and the failure phase was also significant, $F(2, 22) = 12.66$, $p < 0.01$, $\eta_p^2 = 0.365$,

which revealed that participants who encountered a greater number of reliable trials before the first failure (and hence who might have been expected to show more complacency) had a significantly longer response latency than those who experienced fewer reliable trials prior to the FF.

Table 5. 3 Mean Probe Response Latency during the First Failure (Standard error in parenthesis)

	Pre-Failure	Failure	Post-Failure
Early	14.50 (1.38)	14.80 (1.21)	15.00 (2.34)
Late	14.02 (0.72)	31.20 (2.90)	10.60 (0.99)

5.3. Discussion

The current experiment addressed the extent to which an imperfect ATC CRA could assist controller performance, and provided results relative to the more general theory of HAI. The hypotheses and the results are summarized in Table 5.4.

Table 5. 4 The Summary of the Hypotheses Result

Hyp. No	Hypotheses	Results
H1	Imposing the CRA would increase both performance and reduce workload (when compared to fully manual conflict resolution), given that controllers use (depend upon) the automation.	Confirmed
H2	The benefits in (1) would be amplified under high traffic load, when cognitive resources were scarcer.	Confirmed
H3	The benefits of the CRA would be manifest even	Confirmed

	when it was imperfectly reliable, given that its reliability was greater than the threshold of automation assistance between 0.70 & 0.75 (Wickens & Dixon, 2007).	
H4	The CRA, whether perfect or imperfect would decrease situation awareness (SA) of the traffic situation.	Not confirmed
H5	This loss of SA in (4) would be manifest in compromised performance on the occasion of the FF of imperfect automation as compared to manual performance.	Not confirmed
H6	The effects in (4) would be amplified under high workload (high traffic load).	Not confirmed
H7	After automation had failed, the trust in the CRA would decline but the dependence would not decline, due to the high traffic load.	Confirmed
H8	High time pressure should decrease performance, increase situation awareness, and increase workload	Not confirmed

The first hypothesis was confirmed regarding performance and objective workload, replicating the prior findings of Study 1 (H1) but here at the much higher traffic load of twice and three times the volume. Participants used (depended) on the automation, and their maneuver resolution performance with the aid was substantially above unaided performance. Furthermore, workload was reduced by automation, since ATCOs' ability to effectively and safely maintain aircraft separation depends upon the maintenance of their mental picture (Nunes & Mogford, 2003; Alexander, Wickens & Merwin, 2005); the CRA could reduce the amount of mental computation thus reduce their workload. Increased

CRA benefit was also observed in resolving conflicts under higher (47.8%) than under lower traffic load (27.2%), supporting Hypothesis 2.

Regarding Hypothesis 3, the analysis on trial 4, at which time, from the participant's perspective, automation was now seen to be 80% reliable in the two unreliable conditions, no significant decrement in performance was observed relative to the 100% reliable condition, hence providing additional data in support of the conclusions of Wickens & Dixon (2007) that reliability as low as 80% does not impose substantial costs, replicating the effects of study 1 (H2) and thereby generalizing their conclusions from automated detection tasks, to those of automated decision aids.

Furthermore, in the absence of a situation awareness loss in prior monitoring (which, had it occurred, would have been related to the potential build-up of complacency), participants did NOT demonstrate the classic "first failure" effect of automation bias (Clegg, Vieane, Wickens, Gutzwiller, & Sebok, 2014; Dietrich Manzey et al., 2014; Molloy & Parasuraman, 1996; Yeh, Merlo, Wickens, & Brandenburg, 2003) across traffic densities. On the occasion of automation failure, the conflict resolution accuracy was higher relative to the manual performance and no performance difference between the FF occasion and the subsequent failures. These findings suggest that the effect of the first automation failures (Yeh, Merlo, et al., 2003) was not manifest in the current data for performance, indicating that participants were not complacent about the automation. Also, conflict resolution was slightly faster in the "Early" group than in the "Later" group, showing an effect that was opposite to what would predict from the development of complacency, where encountering a greater number of correct automation trials would lead to more rapid acceptance of automation advice. Moreover, SA probe

response latency was much lower during the occasion of FF. This was because participants were occupied with dealing with automation imperfection and set aside the response to the SA probe.

All of these findings suggest that controllers were not “blindly” following automation, but rather exploiting automation and considering the raw radar data more carefully (Xu, Wickens, & Rantanen, 2007). In addition, the workload increased significantly during post-failure. This latter effect is consistent with the assumption that even more attention was paid to the “raw data” regarding conflict resolution after the automation was found to be imperfect, at the expense of attention to other routine tasks. Thus, collectively, these negative (null) effects failed to confirm the “automation dependence syndrome”, embodied by Hypotheses 4, 5, and 6.

The FF findings might be explained as the effects of how the “failure” in the experiment was defined. The failure was not placed on the primary conflict, but the triggering of a secondary conflict due to an acceptance of the CRA, leaving participants to resolve the secondary conflict in 100 seconds or 4 minutes. Hence, upon acceptance, participants might still be able to notice a secondary conflict triggered by his or her decision. This makes the failure to be somewhat unexpected rather than totally unexpected, which is the implication of the research.

The effects of CRA on trust and dependence (H7) were in the expected direction. Trust was degraded in the unreliable automation blocks, replicating the effects observed in study 1 (H5). However, dependence on the CRA was not degraded after it had failed, thus both effects supporting Hypothesis 7. The non-reduced dependence might occur due to

high workload so that participants opted to rely on the CRA to preserve cognitive resources for other ATC tasks although they did not trust it as much.

Regarding H8, the effects of time pressure in this experiment were muted, just as they were in Study 1 (H6). Time pressure did not affect performance and SA. This might be because even the high time pressure occurred within their cognitive capacity and participants might be familiar with this condition thus not compromised their performance and SA. Consistently, low time pressure increased workload. This effect might be because the CRA used in this study served as both conflict alert and resolution aid covering stage 1, 2 and 3 of automation. Consequently, it imposed greater workload when unreliability occurred after participants had shifted their attention away from any potential secondary conflict since the automation assisting information inputs through sensory receptor (Stage 1) and information processing (Stage 2) was also absent. In this condition, the unexpected off-nominal situation contributed to excessive workload due to increased attentional effort for the off-nominal events (Neal et al., 2013; Onnasch, Ruff, et al., 2014).

Chapter 6

Study 3: Integration of a Vertical Situation Display with the Conflict Resolution Aid and their Interaction Effects on ATCO Performance

The aims of Study 3 were to evaluate whether the VSD could support ATCOs performance and to examine how the presence of the VSD would interact with CRA unreliability, whether the presence of a VSD could compensate for any off-nominal event that occurs when the CRA works imperfectly. In contrast with studies 1 and 2, here imperfection was imposed by providing an incorrect resolution for one of the primary conflict encounters. In addition to examining overall ATCO performance with an 80% reliable aid, specific performance on the first (and only) time that the aid failed was also examined. This is because first failure effects often show exaggerated costs imposed by the automation error, so that its imperfect reliability implications are easier to measure (Yeh, et al. 2003; Wickens, Clegg et al. 2015). The independent variables were automation and VSD conditions. There were four dependent variables: overall performance, first-failure performance, situation awareness (because declines in SA have been shown to degrade the human response to failure), and workload. Four hypotheses listed in Chapter 1 were tested.

6.1 Methods

6.1.1 Participants

Twenty (20) ATCOs and four (4) college students (17 males and 7 females) aged 22 to 62 years (Mean = 30 years, SD = 10.29 years) participated in this study. The ATCO participants included controllers from the Civil Aviation Authority of Singapore (CAAS) and Singapore Air Force (SAF). Participants' average work experience was 5.17 years. The student participants were educated in NTU and joined a two-week ATC training conducted by the ATMRI. Informed consent approved by the NTU IRB was provided by each participant.

6.1.2 Apparatus

6.1.2.1 The ATC Simulation Setup

A medium fidelity of ATC simulator, NLR Air Traffic Control Research Simulator (NARSIM), was used to generate various air traffic scenarios. In this study, NARSIM represented the Terminal Radar Control (TRACON) facility of Singapore Airspace and used Standard Instrument Departure (SID) and Standard Arrival Routes (STAR) similar to those used by Singapore Airspace.

The experiment setup consisted of one ATCO position and two pseudo-pilot positions. In the ATCO position, there were four screens, including:

- (i) Two 28.05-inch square format 2K monitors that were used as a primary radar display and a flight data display respectively.
- (ii) A 12.1-inch touch-based monitor that was used to display the CRA.
- (iii) A 22-inch touch-based monitor that was used to display the VSD.

In the pseudo-pilot position, there were three screens, including:

- (i) Two 28.05-inch square format 2K monitors that were used to display the primary radar display and the blipper inputs for each pseudo-pilot.
- (ii) A 12.1-inch touch-based monitor that was used to display the CRA feedback uplinked from the ATCO position.

The primary radar display (Figure 6.1) shows a geographical map of Singapore with the SID and STAR and two active runways. The instrument landing system (ILS) approach procedure was applied. Flight progress strips (FPS) (Figure 6.2) for both arriving and departing aircraft were presented on the right side of the primary radar display. The FPS provided aircraft information and updates.

Figure 6. 1 Primary Radar Display. Blue and green lines represent SID and STAR, respectively. Parallel yellow lines indicate extended runways for interception to the ILS localizer.

WSSS DEPARTURE PENDING										WSSS ARRIVAL SEQUENCE									
20C - 20R SEQUENCE																			
BRO7793	A319	0000	360	30						YMR88	20C								
BA680	8772	2346	320	30						AN68	20C								
TDW2187	A320	2308	300	30						TON18	20C								
AIQ8937	A320	2303	360	30						YMR58	20C								
LM714	8739	2301	340	30						YMR88	20C								
LA0636	A320	2300	360	30						YMR88	20C								
MA8442	8738	2348	360	30						YMR88	20C								
PAL807	A320	2346	360	30						YMR88	20C								
BLK386	A319	2346	320	30						TON18	20C								
JS4492	A320	2343	300	30						MA518	20C								
PAL802	A320	2341	340	30						YMR88	20C								
BLK326	8738	2339	340	30						YMR88	20C								
BLK321	A320	2336	320	30						TON18	20C								
BA874	8772	2336	360	30						YMR88	20C								
BTY106	8739	2334	360	30						AN68	20C								
TDW2184	A320	2303	360	30						YMR88	20C								
BA680	A320	2301	360	30						AN68	20C								
21-OTHER SEQUENCE																			
TMA PENDING										20R - SEQUENCE									
BA311	A333	0015	360	360						20R	0013								
AWR0384	A320	0014	330	330						20R	0015								
AON1724	A320	0011	360	360						20R	0013								
MMA232	8738	0019	330	330						20R	0011								
DOC7069	A319	0009	320	320						20R	0012								
VL0003	A319	0006	320	320						20R	0012								
AWR0381	A320	0001	270	270						20R	0020								
21-OTHER SEQUENCE																			

Figure 6. 2 Flight Data Display. Yellow and blue rows indicate FPS for arriving and departing aircraft respectively.

The pseudo-pilots were equipped with a primary radar display and a blipper on a single monitor to properly understand the airspace situation while directing aircraft. The blipper is a tool used to observe the flight status and to input maneuver commands to an aircraft.

6.1.2.2 The Conflict Resolution Aid

The CRA aid used in this study worked based on the principles used for resolution aircraft and manoeuvre selector (RAMS) (Erzberger, 2006), similar to the one used in Study 1 and Study 2.

6.1.2.3 The Vertical Situation Display

Design Features

The VSD used in this study has five design features. First, color coding for aircraft displayed on the VSD is similar with the ones shown on the primary radar. In particular, the conflicting aircraft pair is shown in red while the other aircraft under an ATCO's control are shown in white. Second, the waypoints along an aircraft's route on the primary radar are also displayed on the VSD. Collectively, these two design features are applied to improve the perceptual-cognitive linkages or "visual momentum" (Olmos, et al. 2000; Woods, 1984) between horizontal and vertical awareness in a manner that adhered to the proximity compatibility principles (Wickens, et al. 2004).

Third, the vertical trend information was provided on the VSD. This is to improve the trend visibility (Wickens, et al. 2013) of vertical information, and not just provide vertical position information.

Fourth, the prediction of vertical trend information is provided comprehensively with the timing information about when aircraft will be at particular route points. This is essential information for controllers to decide on the appropriate resolution maneuver.

Provision of the prediction of trend information can reduce mental computation (Wickens, et al. 2000; Wickens, et al. 1998).

Fifth, ATCOs were given freedom to activate the VSD for the aircraft in their scope of attention, thus they are able to declutter the display (by removing the VSD) if they wish. ATCOs could activate the vertical profile for respective aircraft by clicking on the aircraft's call-sign and re-clicking the call-sign to remove the information. Therefore, they could properly obtain the VSD information without permanently obscuring information regarding other traffic aircraft. This feature satisfies the "detail on demand" principle (Shneiderman & Plaisant, 2005) for the VSD to help ATCOs conserve their attentional resources.

An Application Example

Figure 6.3a shows an example for the situation where there was a predicted conflict between AXM1805 and SLK331. Both aircraft were co-altitude at the cruise phase with trailing aircraft faster than the leading aircraft. Two minutes before the conflict, the aircraft label turned red (as shown in Figure 6.3a), indicating that this conflicting aircraft pair required attention. ATCOs could effortlessly examine this situation by activating the VSD of both aircraft, as shown in Figure 6.3b. ATCOs could also straightforwardly examine whether a secondary conflict would be triggered with the traffic aircraft AWQ8203. In addition, the route points that would be followed by the aircraft according to the flight plan were also provided on the VSD. Both aircraft would follow the route points from ARAMA until they landed on the runway, as indicated by a yellow rhombus (Figure 6.3b). The vertical trend information was also provided on the

VSD as shown in Figure 6.3b. At a glance, it could be seen that both aircraft were aiming to descend for the approach phase.

Next, from the VSD (Figure 6.3b), it could earlier be seen that if ATCOs applied vertical separation, the aircraft would not have time to reach the directed altitude while simultaneously maintaining the lateral separation (i.e. 5 NM), as indicated by the close distance of the green triangles (i.e. the route points). In addition, both aircraft would be nearly laterally parallel at the different altitudes until BOBAG. However, this situation could not be accepted since both aircraft were aiming to approach, thus ATCOs had to make appropriate lateral and/or speed separations for the appropriate landing sequence before BOBAG, which was the initial point for the STAR. Hence, with the aid of VSD in providing the prediction of vertical trend information, ATCOs could also determine appropriate resolution maneuvers. In addition, the VSD shows that there would be no secondary conflict with aircraft AWQ8203 since it would pass BOBAG at different altitude and much earlier than the two conflicting aircraft.

Lastly, in this example, the vertical information was activated for only the three aircraft, AXM1805, SLK331, and AWQ8203 to examine the conflict situation without information clutter from other traffic aircraft.

(a)

(b)

Figure 6. 3 The Plan View (a) and Vertical situation display (b) with the application of the HF principles

6.1.3 Experiment Design and Procedure

The independent variables in this study were automation condition and VSD. The automation condition was a within-subjects factor that included three levels: reliable, unreliable, and manual conditions. The three testing conditions were counterbalanced between participants using a balanced Latin square method. The VSD was a between-subjects factor and was defined by two levels: presence and absence of the VSD. The participants were randomly assigned to these two display conditions, with each condition having twelve participants. The student participants were assigned equally to each condition (i.e. 2 students to each group).

In the reliable condition, the CRA provided correct maneuvering advisories to all potential conflicts (100% CRA aid reliability). In the unreliable condition, there was imperfect advice that provided incorrect resolution to the predicted conflict (80% CRA aid reliability). For example, in Figure 6.4 the CRA advised the aircraft to descend to flight level 140. This advice was inappropriate since the lateral separation could not be met while descending to the instructed flight level. In the manual condition, no CRA was provided; hence, participants performed the ATC tasks manually in this condition. There were five conflict trials in each condition. On the single automation error trial (for the imperfect condition), this error was imposed on the 4th conflict.

Figure 6. 4 Example of Conflict in the Unreliable Condition

The study was a laboratory-based experimental study. Participants were provided with a one-hour pre-experiment session that included briefing and training. ATCOs were given instruction on how to use the CRA and the VSD during the pre-experiment session. During the experiment session, participants communicated with the pseudo-pilots using voice transmission, and controlled all departing and arriving aircraft within their assigned zones. The traffic density was 60 aircraft (i.e. 2X of that in Study 1). Participants are required to (1) accept incoming aircraft and hand over aircraft leaving their controlled area (2) provide appropriate clearance for altitude and speed according to the SID and STAR, (3) assign appropriate approach clearance to intercept ILS localizer, (4) maintain separation between aircraft and resolve air traffic conflicts. Upon receiving voice instructions from the controllers, the pseudo-pilot provided the respective commands to the ATC simulator.

Participants performed the monitoring and controlling tasks with the aid of the VSD and the CRA. While performing the tasks, participants could make use of the VSD to supply aircraft vertical information and prediction as elaborated in Section 6.2.2.3.

Participants could activate aircraft's vertical information by clicking on the respective aircraft's call-sign. Upon clicking the call-sign, the aircraft's vertical information would appear on the VSD. If the aircraft's information was no longer necessary, participants could remove the vertical information on the VSD by re-clicking on the aircraft's call-sign.

Five pre-set conflicts were placed in each scenario corresponding to the automation conditions. The sector density was 60 aircraft. The CRA provided a resolution advice two minutes prior to a conflict (i.e., loss of separation). Participants could either accept or reject the resolution advice provided by the CRA by clicking either the accept or the reject button. If the ATCOs accepted the advice, the resolution advice would be automatically uplinked to the pseudo-pilot's screen. The pseudo-pilot would directly apply the resolution by executing preprogrammed commands in the simulator. If the ATCOs rejected the resolution advice, the CRA would stop processing the respective aircraft' data, thus ATCOs needed to provide their own resolution maneuver that would be executed by the pseudo-pilots. Participants were also required to respond to SA probes that appeared in every 6 minutes throughout a one-hour experiment run. The SA probes were pertaining to the conflicting aircraft as well as other aircraft in the airspace. SA probes were presented using the Situation Present Assessment Method and were elaborated in details in Chapter 3.

6.1.4 Dependent Measures

The dependent measures were task performance, mental workload, and situation awareness. The details regarding the dependent measures are presented in Chapter 3.

6.1.5 Analysis

The conflict resolution time was only collected in the two CRA conditions because, in the manual condition, there was no clearly defined event to start a timing clock. Therefore, those data were analyzed using a 2(automation level) x 2(display) mixed-design ANOVA (analysis of variance). The remaining dependent variables (i.e., percentage of resolved conflicts, mental workload, and situation awareness measures) were collected in the three CRA conditions (i.e. reliable, unreliable, and manual conditions). The data were analyzed using a 3(automation level) x 2(display) mixed-design ANOVA. Prior to the analyses, ANOVA assumptions were tested as explained in Study 2. Furthermore, t-test analysis was also conducted for significant interaction effects as appropriate to systematically see the effects of VSD on different automation levels.

Also, analyses of the failure trial, and the four non-failure trials were performed for the conflict resolution performance in both display conditions. The correct trials data reflected the collected mean of all non-failure trials. These analyses were performed to see the effects of the CRA imperfection on the failure occasion.

6.2 Results

6.2.1 Task performance measures

An ANOVA revealed a significant main effect of automation type ($F(1.55, 33.99) = 3.96, p = 0.038, \eta_p^2 = 0.153$), and no significant effect of VSD on conflict resolution performance ($F(1, 22) = 1.19, p = 0.29, \eta_p^2 = 0.051$). As shown in Figure 6.5 and Table 6.1, post hoc analyses revealed that reliable automation ($M = 91.67\%$, $SE = 4.44\%$)

supported better performance than the manual condition ($M = 83.34\%$, $SE = 5.38\%$) ($p = 0.039$). Better performance was also observed when participants were equipped with imperfect automation ($M = 93.34\%$, $SE = 3.61\%$) than manually performing the task ($M = 83.34\%$, $SE = 5.38\%$) ($p = 0.044$). Performance with the two automated conditions ($M = 91.67\%$, $SE = 4.44\%$ vs $M = 93.34\%$, $SE = 3.61\%$) did not differ from each other ($p = 1.000$). Automation level and VSD did not interact ($F(1.55, 33.99) = 0.05$, $p = 0.95$, $\eta_p^2 = 0.002$).

Figure 6. 5 Conflict Resolution Performance (error bars indicate ± 1 SE)

Table 6. 1 Mean Percentage of Resolved Conflict (SE in parentheses)

	Reliable	Unreliable	Manual
VSD	95.00% (2.61%)	96.67% (3.33%)	85.00% (5.00%)
No VSD	88.33% (6.26%)	90.00% (3.89%)	81.67% (5.75%)

Table 6. 2 Mean Conflict Resolution Time (SE in parentheses)

	Reliable	Unreliable
VSD	20.67 (2.58)	21.33 (3.45)
No VSD	17.20 (2.68)	16.90 (2.01)

For conflict resolution time (Table 6.2), there was no significant effect of display ($F(1, 22) = 1.42, p = 0.25, \eta_p^2 = 0.061$) nor automation level ($F(1, 22) = 0.09, p = 0.93, \eta_p^2 = 0.000$). The interaction between automation level and VSD was not significant ($F(1, 22) = 0.06, p = 0.81, \eta_p^2 = 0.003$).

6.2.2 Mental Workload

Participants rated their workload lower with the VSD than without it ($F(1, 21) = 10.54, p = <0.01, \eta_p^2 = 0.334$) (Table 6.3). No significant interaction effect between automation level and VSD was found ($F(2, 42) = 0.12, p = 0.89, \eta_p^2 = 0.006$). The workload rating across automation conditions was not different ($F(2, 42) = 0.64, p = 0.53, \eta_p^2 = 0.030$).

Table 6. 3 Mean Workload Rating (SE in parentheses)

	Reliable	Unreliable	Manual
VSD	60.56 (2.90)	59.99 (3.09)	64.05 (2.63)
No VSD	73.94 (3.36)	74.00 (5.07)	75.51 (4.22)

Consistent with the workload ratings, a trend of lower objective workload was found with the VSD as indicated by lower percentage of timeouts ($F(1, 19) = 2.86, p = 0.10, \eta_p^2 = 0.131$) (Table 6.4). The interaction effect between automation level and VSD was not significant ($F(2, 38) = 1.73, p = 0.19, \eta_p^2 = 0.083$).

The effect of automation level on the objective workload measure was significant ($F(2, 38) = 4.61, p = 0.02, \eta_p^2 = 0.195$). The post-hoc analyses revealed that the reliable CRA ($M = 11.88\%$, $SE = 3.98\%$) greatly reduced the workload as compared to manual performance ($M = 22.52\%$, $SE = 6.13\%$) by lowering the timeouts ($p < 0.01$). There was no difference in the overall objective workload between reliable ($M = 11.88\%$, $SE = 3.98\%$) and unreliable conditions ($M = 15.14\%$, $SE = 4.79\%$) ($p = 0.30$) as well as in the 8th probe timeout ($M = 33.33\%$ vs $M = 28.57\%$, $t(19) = 0.36, p = 0.72, r = 0.05$), suggesting the benefit of the CRA.

Table 6. 4 Mean Percentage of Timeouts (SE in parentheses)

	Reliable	Unreliable	Manual
VSD	3.50% (1.26%)	9.00% (3.66%)	20.60% (6.38%)
No VSD	20.26% (6.69%)	21.28% (5.91%)	24.44% (5.88%)

6.2.3 Situation Awareness

For situation awareness accuracy, there was a strong benefit for the VSD ($F(1, 20) = 5.77, p = 0.03, \eta_p^2 = 0.224$), and for the CRA ($F(2, 40) = 14.39, p < 0.01, \eta_p^2 = 0.419$) (Table 6.5). However both of these were modified by a marginally significant interaction

($F(2, 40) = 2.95$, $p = 0.06$, $\eta_p^2 = 0.128$), which can be described by the large benefit of the VSD when automation was reliable ($t(20) = 3.06$, $p = 0.01$, $r = 0.49$), a modest benefit when automation was imperfect ($t(20) = 2.47$, $p = 0.03$, $r = 0.54$) and no benefit at all in the manual condition ($t(20) = 0.95$, $p = 0.37$, $r = 0.15$) (Figure 6.6).

Table 6. 5 Mean Percentage of Correct SA Probes (SE in parentheses)

	Reliable	Unreliable	Manual
VSD	87.00%	66.00%	56.73%
	(3.26%)	(3.13%)	(4.44%)
No VSD	59.64%	46.55%	49.05%
	(9.06%)	(5.44%)	(9.29%)

Figure 6. 6. Interaction effect between automation conditions and display on SA probes accuracy (error bars indicate ± 1 SE)

While the effects of automation on probe reaction time (RT) were significant ($F(2, 36) = 4.04$, $p = 0.03$, $\eta_p^2 = 0.183$), these can only be interpreted in the context of the significant interaction between automation level and VSD, revealing that all RTs were

statistically equal except for an extremely short RT to SA questions when automation was reliable and there was no VSD. This interaction is not readily interpretable.

Table 6. 6 Mean Probe Response Latency (SE in parentheses)

	Reliable	Unreliable	Manual
VSD	15.08 (1.77)	16.50 (2.12)	14.17 (1.59)
No VSD	9.82 (0.89)	17.08 (2.13)	16.64 (2.23)

6.2.4 First failure effects

The first-failure effects were analyzed to investigate the complacency that might occur during the ATC tasks performance with the CRA. The presence of first-failure effects might also reveal the benefits of VSD to off-set the CRA imperfection.

Table 6.7 presents conflict resolution accuracy as a joint function of the all non-failure trials versus the failure trial and the presence or absence of the VSD. The table shows the very modest (6%) non-significant loss of performance upon automation failure when the VSD was present ($z = 1.08$, $p = 0.28$, $r = 0.14$), in contrast to the large 29% loss in the absence of the VSD ($z = 3.01$, $p < 0.01$, $r = 0.39$).

Table 6. 7 Conflict Resolution Accuracy for the Correct and Failure Trials

	Correct Trials	Failure Trial
VSD	97.92%	91.67%
No VSD	95.83%	66.67%

6.3 Discussion

The purpose of this study was to replicate the benefits found previously for a CRA, even as those benefits may have been reduced when the aid was imperfect; to examine the independent benefits of a VSD and for conflict resolution; and to establish if such an aid provided a particular benefit upon the failure of the imperfect aid, because of the VSD support for SA, a critical element in aiding the recovery from automation failures. These goals were addressed in the four primary hypotheses of the study.

H1 was confirmed, replicating the two prior studies. The benefit of the aid, relative to manual performance was small (93 vs. 83%) but significant. However the smaller magnitude of the benefit, relative to Study 1 and Study 2, was partially because the subjects in the manual condition performed substantially better here than in the previous study. This could be explained by the fact that most of the participants in the current study were ATCOs, as compared to the ATC student participants in the previous studies (Study 1 and Study 2). Most prominently, however, the CRA was found to improve the students' performance to the similar accuracy level as the ATCO participants. Importantly, even when CRA reliability was reduced to 80%, overall conflict resolution performance was still better than in the manual condition, consistent with both prior studies in this dissertation, and with findings regarding automated detection aids (Wickens & Dixon, 2007).

In addition, supporting the H1, both subjective and objective workload were also found lower with the CRA as compared to manual performance, and the imperfection of the CRA was not found to increase the objective workload, suggesting the benefits of the CRA, regardless of its unreliability. Since ATCOs' ability to effectively and safely

maintain aircraft separation is influenced by their mental computation (Alexander, et al. 2005), the CRA could reduce the amount of mental computation, thus reducing their objective workload.

Regarding H2, a small, non-significant cost to overall performance accuracy in the 80% reliable condition was indeed found. Overall however the imperfect automation performance deficit was diluted because, in the first three conflicts in that condition, the CRA did not fail, and so its performance on those trials was indistinguishable from that of perfect automaton. However when data analysis focused just on the trial in which automation failed (Trial 4), the results revealed a performance decrement relative to the prior perfect trials (Table 6.7), and this decrement was particularly pronounced when the VSD was absent, as discussed below.

The benefits of the VSD, proposed in H3, were observed in the improved SA and lower workload, but not in overall accuracy (i.e., across all trials). The VSD was found to support the maintenance of the ATCO's mental picture (Nunes & Mogford, 2003) since it provides updates for altitude trends and changes that are not available on the primary radar display, thus lowering their mental workload.

Associated with the workload reduction, the presence of VSD improved the SA as indicated by the 18% improvement in SA probe accuracy. This was arguably because the VSD helped ATCOs maintain their mental picture about aircraft's states, thus preserving the attentional resources to attend to other tasks. This finding was in contrast to that of Marusich, et. al. (2016) who found SA reduction with increased task-relevant information in command and control in the military context. However, Marusich, et. al. measured the

SA using a self-reported subjective measure that might not well reflect actual SA as people are typically not good at reporting information regarding mental events (Endsley, 1995; Marusich, et al. 2016; Wickens, 2000). In addition, the VSD design principles in this study including (i) improvements in the perceptual-cognitive “linkages” (ii) improvements in the trend visibility of vertical information, (iii) providing the prediction of trend information, and (iv) allowance of the “detail on demand” principle (please see Section 2.2.3) were found to positively affect SA in a way that was not done by Marusich et al. In Marusich et al.’ study, the task-relevant information and the map of operation area was not quite cognitively linked so that participants still needed to carefully scan back the area of operation when the information is available.

Despite the null effect of VSD in the overall conflict resolution accuracy, the accuracy effects **were** observed in the more focused analysis on the failure trial, addressed in H4. The hypothesis asserted that the costs of automation failure could be muted or offset, by the benefits of increased SA fostered by the VSD, consistent with the findings of the meta-analysis of Onnasch et al. (2014). This differential effect was indeed revealed in the interaction depicted in Table 6.7, with the VSD restoring imperfect automation performance to a level (80%) approaching, although not quite equal to that of the perfect automation observed in the first three trials. Thus these results provide another data point to support the idea that the transparency of visual displays can offset the costs of imperfection in automation (Mercado et al., 2016; Hoff & Bashir, 2015; Rovira et al., 2014).

In the present study, the VSD that provides transparency to automation recommendations would mitigate the cost of automation imperfection. But the finding,

particularly for resolution performance, was weak as it appeared to be true only in response to the first failure, and not in overall resolution performance. In order to enhance the application of the VSD in supporting ATCOs' resolution performance, further research examining different faulty conditions in a longer experiment run with longer training period is required.

In terms of the improvement for the current VSD design, there are three possible design features that may improve ATCOs performance variables. First, a highlighting option can be provided. This feature will allow ATCOs to watch an aircraft being highlighted on the radar display when they click on the aircraft on the VSD, or vice versa. Second, the VSD should provide an option to depict the CRA's recommendation before it is even applied by the ATCOs. This feature is expected to lower the performance decrement and to better increase the SA. Third, the information of climb rate or descend rate may be provided along with the vertical trend line so that it will be easier for ATCOs to obtain the details of the graphical information.

No matter whether it is reliable or unreliable, the CRA improved SA accuracy with the presence of VSD than without it. This was because the CRA helped reduce the mental effort for conflict resolution resulting in more attentional resources available to monitor the airspace (Murphy, et al. 2012). Moreover, the participants were further assisted in monitoring the airspace situation and could observe the CRA's decision better with the VSD.

One interesting finding of the current study is that on the failure trial, resolution performance of the imperfect condition without the VSD (67%) was at a level well below

that of the unaided manual condition (82%), thereby suggesting the presence of the well-established complacency effect (Mosier, et al. 1998; Wickens, et al. 2015). One would assume that a greater amount of exposure to occasional automation failure trials would enable the controller to better calibrate monitoring of the raw data.

Chapter 7

Conclusions

7.1. Key Findings and Contributions

This research focused on the important automation and environmental factors that may influence the use of the CRA, and experimentally examined the effects of these factors on ATCOs' performance variables in three separate studies. Key findings and contributions are summarized as follows:

Study 1

Existing studies regarding the CRA (Cardosi, et al, 1992; Dwyer & Landry, 2009; Prevot, Homola, & Mercer, 2008; 2012; Jha, et al. 2011; Borst, Westin & Hilburn, 2012; 2013; Cabrall, et al., 2014; Goritzlehner, et al. 2014) have not investigated the effects of automation reliability, which is indisputably one of the most important aspects of HAI. In addition, no prior study has addressed the issue of time pressure in conflict resolution although it has been addressed in conflict detection (Eyferth, et al. 2003; Metzger & Parasuraman, 2005). Study 1 made scientific contribution to the Human Factors and Ergonomics field by it revealing the mechanisms through which underlying factors of conflict resolution safety affect ATCOs' performance variables, and extends the findings of imperfect reliability benefits beyond detection systems (Wickens & Dixon, 2007) to decision aiding.

Regarding the effects of automation reliability, according to previous research in conflict detection and in other command and control domains (Parasuraman, et al. 1993;

Bailey & Scerbo, 2007; Rovira, et al. 2007; 2010; Rice, 2009; Wang, 2009; Chen & Barnes, 2012), automation imperfection would lead to a detrimental effect. However, the current study found that the availability of CRA was associated with improved performance, reduced workload, and increased situation awareness, regardless of its imperfection, above the manual performance. These findings suggest the importance of the CRA and its practical implications in ATC operations. In particular, system designers should set a safety layer of automation for conflict resolution, and they should place greater effort in providing reliable automation that assists in the first conflict resolution iteration. This shifts the complexity of the CRA development to a feasible level. As of today, research is still ongoing for the development of CRA; however, considering every single constraint in the airspace to provide a fully secondary conflict-free resolution is nearly impossible (Kuchar & Yang, 2000). Yet, ATCOs are waiting for the real support in facing the imminent traffic growth.

Study 1 also revealed that time pressure did not raise safety concerns in the conflict resolution process. No detrimental costs on performance, workload, and situation awareness were observed in high time pressure conditions, although it is quite likely that more excessive time pressure might have degraded resolution performance as well as increasing dependence. This finding is consistent with that in conflict detection (Metzger & Parasuraman, 2005). This suggests, in addition to the first implication, that designers should set a time allowance to any potentially triggered secondary conflict up to two minutes look-ahead-time (LAT), in a similar fashion to that in the existing conflict detection system.

Study 2

The findings from Study 1 provided strong quantitative evidence for the benefits of the CRA, regardless of its imperfection. As stated earlier, ATC is facing the major challenge of an increase in air traffic density. The knowledge of whether the CRA is applicable in high traffic environments has become increasingly relevant and has not been addressed in the existing literature (Galster, et al. 2001; Prevot, et al., 2009; 2012). The advantage of Study 2 is that it examined the interaction between the CRA reliability and traffic density on ATCOs' performance variables.

The results showed that no matter whether the CRA is reliable or imperfectly reliable, the addition of the CRA would improve ATCOs' performance and reduce their workload in both lower and higher traffic densities, thereby replicating a key result of Study 1. In addition, consistent with prior studies in conflict detection (Galster, et al. 2001; Metzger & Parasuraman, 2001; 2005; 2006; Prevot, et al. 2012) better conflict resolution performance was found in lower traffic, however, greater benefits of the CRA were found as traffic density increased. These findings provide empirical evidence regarding the need for the provision of the CRA in high traffic density environments in future ATC operations. However, the reliability level should be assured to be above the threshold of automation assistance which is around 72% (Wickens & Dixon, 2007).

In addition, ATCOs were not complacent with the CRA, indicating that they worked well with the CRA and preferred to share the responsibility of separation assurance, consistent with the past study by Cabrall, et al. (2014). Thus, setting automation at the moderate level (Bekier, Molesworth, & Williamson, 2012; Li, et. al, 2014; Parasuraman, et al., 2000) that supports rather than replaces ATCOs (Prevot, et al.

2012) was recommended in developing the CRA as how it was applied in this Study. This allows ATCOs to complement the CRA safely with their manual intervention, should any automation failure occurs.

Study 2, however, showed that reduced situation awareness was manifest during the first-failure occasion and increased workload was observed during the post-failure occasion. Although no detrimental effects on performance were observed during the first-failure showing that complacency effect was absent, in contrast to a prior study in process control (Clegg, et al, 2014). These effects could be addressed by appropriate training (Higham, Vu, Miles, Strybel, & Battiste, 2013) prior to the use of the CRA, as well as by providing a support to rapidly obtain clear inferences of airspace information in order to maintain SA (Marusich, et al. 2016) and workload during and after the failure occasion. The latter implications has been applied and examined in the Study 3.

Study 3

For pilots, a lack of understanding about the surrounding airspace has been well addressed with the conflict detection and traffic information (CDTI) that incorporates horizontal and vertical situation display (Thomas & Wickens, 2005). However, in ATC facilities, no vertical situation display has been provided thus far. Theoretically, the advantage could be transferred to the ATC facilities for ATCOs who also require a swift understanding of airspace inferences. However, no empirical evidence regarding the benefit of vertical situation display for ATCOs has been available. Study 3 was conducted to investigate the effects of vertical situation display and its interaction with the imperfection of the conflict resolution automation on ATCOs' performance variables.

The integration of the VSD into ATC facilities has positive benefits on ATCOs' performance. The benefits of the VSD suggest a promising technology for assisting ATCOs in their performance, consistent with the previous studies of VSD for assisting pilot (Heylen, et al. 2008; Prevot & Palmer, 2000). The reductions in workload and improvement in situation awareness provided by the display were pronounced, even if it did not improve direct performance overall. Probably, the greatest benefit was provided by presenting graphic vertical trend information, not just in direction (increasing, decreasing, level), but in the *rate* of altitude change. This benefit parallels to that found by continuous predictive display visualizations in the process control industry (Yin, Wickens, Helander, & Laberge, 2015).

Next, the results of Study 3 showed that providing the vertical situation display when ATCOs work with the CRA did help offset the automation imperfection and reduce performance decrement due to the imperfection. This finding is consistent with a prior study in command and control (Rovira, et al. 2014) and thus extends it to the ATC conflict resolution context. This study suggests the integration of the CRA and vertical situation display in future ATC workplaces to compensate for the CRA's imperfection by supplying complete inferences of airspace information and supporting the awareness of the CRA's accuracy, to make ATCOs remain in the loop (Wickens, et al. 2013).

Collectively, this research comprising of the three studies has major implications for air traffic management, as well as for the science of human-automation interaction. The benefits of the CRA, even if imperfect, continue to suggest that this is a promising form of automation support tool for the ATCO (Prevot et. al., 2012; Martin & Imbert,

2013; Trapsilawati et al., 2015; 2016), particularly if the CRA can avail more resources for the controller to monitor the raw data on the display. There was little evidence for whether such a tool leads to complacency in real world operations, and hence leads to a compromised response to automation failures or imperfections (Onnasch et al., 2014). The mild cost of imperfect automaton on the first failure was balanced by the overall benefit of the CRA, even an imperfect one, to improved situation awareness. However, we would argue that this benefit would only be realized: (1) as long as the raw data of the traffic picture are readily available to the controller (and here, the benefits were enhanced when more raw data were provided by the VSD) and (2) the automation of decision aiding was at a lower level, such that automation only recommends a maneuver, but does not automatically implement it (Parasuraman, Sheridan & Wickens, 2000; Onnasch et al., 2014).

7.2. Lessons Learned Across the Three Studies

Across the three studies, it was found that the CRA always improved participants' performance relative to manual performance, even if it was imperfect. This was supported by the large overall effect size ($f = 0.96$) (Cohen, 1992). The large effect sizes were also consistent across experiments as shown in Figure 7.1. This finding suggests that the CRA is a promising concept for the future ATC that is able to support a wide range of ATCOs' experience and expertise, despite the fact that it may not always be able to offer the optimal resolution advice. Also, there were no adverse consequences on performance due to time pressure according to the replicated results in Study 1 and Study 2, although a

more stringent level of time pressure should be explored in future studies. Regarding workload, the three studies revealed replicated results that the unreliable CRA did not compromise workload. For the situation awareness, it was consistently revealed in the three studies that the participants' situation awareness was not reduced with the presence of the CRA.

While reliable CRA improved conflict resolution performance relative to the unreliable CRA in Study 1, no significant performance differences were found between the reliable and the unreliable CRA in Study 2 and Study 3. Although the reason for this difference remains unclear, it might be related with the different workload imposed in the three studies through the different levels of traffic density. Participants might be less vigilant when they thought that the situation was less complex and still under control (Warm, Parasuraman, & Matthews, 2008) as in Study 1, thus inducing the reduction of active responding and engagement with the CRA, particularly when it was imperfect.

Furthermore, in spite of the fact that the CRA benefits human operators in the three experiments, the benefit was varied across the experiments. In particular, in Study 1 and Study 2 the benefits of the CRA were very large (i.e. about 30%) whereas in Study 3 it was quite modest (i.e. about 10%). However, as shown in Figure 7.1, the overall effect size across the three studies was large. A possible reason for the difference in automation benefit between experiments is the difference in population sample. Study 1 and Study 2 used mostly (90%) students while Study 3 used mostly (90%) ATC professionals. It is logical that professionals benefit less from automation because their baseline (manual) performance was found to be (as expected) much higher.

Figure 7. 1 Forrest plot analysis for performance (overall effect size of the automation condition across the three studies is shown)

In addition, the workload was not different across automation conditions in Study 1, although the other two studies (i.e. Study 2 and Study 3) replicated the finding that workload was lower with the CRA than manually performing the tasks. The overall effect size was also large for the workload with $f = 0.44$ (Figure 7.2), well above the f of 0.40 which is the lower limit for large effect as defined by Cohen (1992). However, it is possible that this workload difference could be associated with the traffic load difference. Since the CRA supports the conflict resolution, its availability in low traffic density, might be associated with “workload insensitivity” where ATCOs still needed to attend to other controlling tasks. However, in higher traffic density, due to a lot of resources needed for controlling many more aircraft, a support in conflict resolution task would seem to be much appreciated by the ATCO.

Figure 7. 2 Forrest plot analysis for workload (overall effect size of the automation condition across the three studies is shown)

It is important to highlight that the replicated finding across the three studies concerning SA and off-nominal performance did not confirm the general findings of the HAI literature, particularly in the relation between situation awareness accuracy and assumed automation dependency. This finding is supported by the large meta-analytic effect size of the automation condition on SA ($f = 0.47$) (Cohen, 1992) that provides a more precise estimate of effect size from the three studies (Figure 7.3). The general findings of HAI, as presented in a meta-analysis (Onnasch, et al. 2014), revealed that more automation could lead to degraded situation awareness and reduced active engagement with the systems, thus compromising the performance when the automation failed. However, the current series of studies in this dissertation revealed that situation awareness was not decreased and performance was not always adversely affected. It could be observed that the participants remained closely attentive to the raw traffic data. It might be explained from two points of view. First, in general, it could be seen as the result of the degree of automation of the CRA, although it is indeed at a later automation stage but it is positioned at a medium automation level (Sebok & Wickens, 2016), where the

participants have final authority over the automation (i.e. either decide to reject or accept the automation).

Another possibility relates to the possible trade-off between situation awareness and workload. In Onnasch et al.(2014), it had been assumed that participants who show better performance with automation are allowing automation to execute that performance, becoming somewhat complacent and failing to monitor automation (hence losing SA). However it may be that, participants take advantage of resources released by automation's decision advising, to allow them to monitor the process more closely, and hence actually improve SA, as appears to be the case in Study 3. Indeed, this possible trade-off between performance assistance of automation and SA seems to be revealed across the three studies here, if the two variables are compared against each other: when there was lower performance gain (Study 3), SA was high. When there was the largest performance gain (Study 2), SA was lowest and for Study 1, the two variables were in the middle.

Figure 7. 3 Forrest plot analysis for SA (overall effect size of the automation condition across the three studies is shown)

In sum, the series of the studies in this dissertation provides replicated results to confirm the benefits of the CRA, regardless its possible imperfection, as an additional

safety layer in future ATC. Moreover, it is worth noting that the presence of the CRA was not observed to compromise ATC safety as how other automation aids might have compromised safety in other contexts. Hence, the application of the CRA for the future ATC is encouraged although it deserves continuing investigations.

7.3. Limitations and Future Directions

There are some limitations in the experimental studies that should be taken into consideration when generalizing the findings. First of all, the air traffic environment simulated in this study might not completely reflect the real situation. Some environmental factors, such as weather, were not taken into account in the study. However, it must be noted that ATC simulation has been used in prior studies and generated suitable outcomes since it is nearly impossible to test a new future-functionality concept that is safety critical using the real-world system (Rovira & Parasuraman, 2010; Bekier, 2012). Still, the study of conflict resolution, even using ATC simulation, has not received much attention and certainly more empirical studies are needed prior to the real application of the novel CRA on ATC facilities.

Second, in Study 1, 2, and 3, the simulation tasks for all conditions including reliable, unreliable, and manual were not conducted in a long experimental trial (e.g. in months per participant). This implied that the findings did not consider the issue of vigilance that may arise in long-term operations. Vigilance has played an increasingly important role as increased automation shifts ATCOs into more managerial positions (Langan-Fox, et al. 2009). Although the main goal of the increased automation in ATC systems is to lower mental workload and enhance SA as well as vigilance, this may lead

to an opposite effect of higher workload. For example, if the CRA turns the role of ATCOs to be more tedious, the vigilance may deteriorate and affect performance during an off-nominal situation. A future research should be conducted for examining the effects of CRA reliability on ATCOs' performance variables in a longer experimental trial to see the effects of the automation imperfection on their vigilance.

Third, in Study 1 and 2, the effects of time pressure were not significant in affecting most of the ATCOs' performance variables. This factor may not have been manipulated within the range in which it had much of an effect. In future research, time pressure may be manipulated in a shorter time range, although it should be longer than the TCAS activation time (Thomas, Wickens, & Rantanen, 2005).

Fourth, in Study 2, the effects of traffic density were only examined with the manipulations of aircraft number. However, ATC complexity is also increased with the traffic increase and is not only determined by aircraft count, but also airspace factors and operational constraints (Histon, et al. 2002). For example, the secondary conflict was preset and provided to appear in one of the conflict trials in Study 1 and Study 2. However, it is logical that conflict resolution should be more difficult in heavy traffic since the resolution advisory may have higher chance to trigger a secondary conflict due to the airspace complexity that depends upon the structure of airspace flow (Sridhar, et al. 1998). Therefore, future research should address the parameter of secondary conflict chance on ATCOs' performance variables.

Fifth, future research should investigate the training of using VSD for ATCOs. In particular, the current ATCOs' tasks are mainly performed using the plan-view radar display and they are not used to the proposed VSD. Although Study 3 has proven that

VSD could lower ATCOs' workload and enhance their SA, but no direct benefit on task performance was observed. This could probably be explained in the way that they required more training to gain larger benefits from the VSD.

Lastly, not all the participants in this study are ATCOs. Hence, the result might slightly bias toward positive responses. However, all the participants had ATC training and knowledge and were familiar with ATC system through prior training. Further, both ATCOs and students have been found in previous study (Rantanen & Nunes, 2005) to have similar preference in resolution maneuvering instruction patterns. Even, in the study by Borst, Westin, & Hilburn (2012), they found that students outperformed ATCOs in all traffic conditions. In all scenarios, students had no LOS whereas ATCOs created LOS. Students were also found to have higher control efficiency than ATCOs. In addition, the results in this research revealed no performance difference between ATCO and student participants showing that the CRA could support a wide range of expertise. Nevertheless, novices also tended to respond conventionally and more rapidly in accordance to conflict definition (Stankovic, Raufaste, & Averty, 2008). This study would then provide a conservative estimate of the ATC practice in using automation for conflict resolution. Since the CRA has not been applied in the current ATC workplace, this research was conducted to gather preliminary information in the CRA context. Using students as participants also well fitted with the explanatory nature of the present study (Goritzlehner et al., 2014). However, more empirical studies examining the proposed CRA using active ATCOs should be conducted in the future.

References

- Airbus. (2013). Future Journey 2013–2032, Airbus Global Market Forecast Report. http://www.airbus.com/company/market/forecast/?eID%4dam_frontend_push&docID%433752.
- Albrecht, Georg H, Lee, Hak Tae, & Pang, Alex. (2012). Visual Analysis of Air Traffic Data Using Aircraft Density and Conflict Probability.
- Alexander, Amy L, Wickens, C. D., & Merwin, David H. (2005). Perspective and coplanar cockpit displays of traffic information: Implications for maneuver choice, flight safety, and mental workload. *The International Journal of Aviation Psychology*, 15(1), 1-21.
- Athènes, Sylvie, Averty, Philippe, Puechmorel, Stephane, Delahaye, Daniel, & Collet, Christian. (2002). *ATC complexity and controller workload: Trying to bridge the gap*. Paper presented at the Proceedings of the International Conference on HCI in Aeronautics.
- Atoyan, Hasmik, Duquet, Jean-Rémi, & Robert, Jean-Marc. (2006). *Trust in new decision aid systems*. Paper presented at the Proceedings of the 18th International Conference of the Association Francophone d'Interaction Homme-Machine.
- Averty, Philippe, Collet, Christian, Dittmar, Andre, Athenes, Sylvie, & Vernet-Maury, Evelyne. (2004). Mental workload in air traffic control: an index constructed from field tests. *Aviation, space, and environmental medicine*, 75(4), 333-341.
- Azuma, Ronald, Daily, Mike, & Krozel, Jimmy. (1996). *Advanced human-computer interfaces for air traffic management and simulation*. Paper presented at the Proceedings of 1996 AIAA Flight Simulation Technologies Conference.
- Bacon, L Paige, & Strybel, Thomas Z. (2013). Assessment of the validity and intrusiveness of online-probe questions for situation awareness in a simulated air-traffic-management task with student air-traffic controllers. *Safety science*, 56, 89-95.
- Bahner, J Elin, Hüper, Anke-Dorothea, & Manzey, Dietrich. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, 66(9), 688-699.

- Bailey, Nathan R, & Scerbo, Mark W. (2007). Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, 8(4), 321-348.
- Bekier, Marek, Molesworth, Brett RC, & Williamson, Ann. (2012). Tipping point: The narrow path between automation acceptance and rejection in air traffic management. *Safety science*, 50(2), 259-265.
- Billings, Charles E, & Reynard, William D. (1984). Human factors in aircraft incidents: results of a 7-year study. *Aviation, space, and environmental medicine*, 55(10), 960-965.
- Bisantz, Ann M, & Seong, Younho. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics*, 28(2), 85-97.
- Boag, Christine, Neal, Andrew, Loft, Shayne, & Halford, Graeme S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics*, 49(14), 1508-1526.
- Boeing. (2015). Statistical Summary of Commercial Jet Airplane Accidents. from Aviation Safety Boeing Commercial Airplanes http://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf
- Brudnicki, DJ, & McFarland, AL. (1997). User Request Evaluation Tool (URET) conflict probe performance and benefits assessment. *MITRE CAASD, MP, 1130065520*.
- Cabrall, CD, Gomez, AN, Homola, JR, Hunt, SM, Martin, L, Mercer, J, & Prevot, T. (2014). Transitioning Resolution Responsibility between the Controller and Automation Team in Simulated NextGen Separation Assurance *Air Traffic Management and Systems* (pp. 147-172): Springer.
- CANSO. (2012). Accelerating Air Traffic Management Efficiency: A Call to Industry *Paper of ATM Global Environment Efficiency Goals for 2050*: Civil Air Navigation Service Organization.
- Canton, Riva, Refai, Mohammad, Johnson, Walter W, & Battiste, Vernol. (2005). *Development and integration of human-centered conflict detection and resolution*

- tools for airborne autonomous operations*. Paper presented at the Proc. 12th International Symposium on Aviation Psychology, Oklahoma City, OK.
- Cardosi, Kim M, & Boole, Pamela W. (1991). Analysis of pilot response time to time-critical air traffic control calls: DTIC Document.
- Cardosi, Kim M, Buerki-Cohen, Judith, Boole, Pamela W, Hourihan, Jennifer, Mengert, Peter, & Disario, Robert. (1992). Controller Response to Conflict Resolution Advisory: DTIC Document.
- Carroll, John M, Olson, Judith Reitman, & Anderson, Nancy S. (1987). *Mental models in human-computer interaction: Research issues about what the user of software knows*: National Academies.
- Chen, Jessie YC, & Barnes, Michael J. (2012). Supervisory Control of Multiple Robots Effects of Imperfect Automation and Individual Differences. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(2), 157-174.
- Clegg, Benjamin A, Vieane, Alex Z, Wickens, Christopher D, Gutzwiller, Robert S, & Sebok, Angelia L. (2014). *The effects of automation-induced complacency on fault diagnosis and management performance in process control*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Cohen, Jacob. (1992). A power primer. *Psychological bulletin*, 112(1), 155.
- Corker, Kevin M. (1999). *Human performance simulation in the analysis of advanced air traffic management*. Paper presented at the Proceedings of the 31st conference on Winter simulation: Simulation---a bridge to the future-Volume 1.
- Craig, Dan, Morales, Dan, & Oliver, Mike. (2004). USS Vincennes incident: MIT Aeronautics & Astronautics.
- Dao, Arik-Quang V, Brandt, Summer L, Bacon, L Paige, Kraut, Joshua M, Nguyen, Jimmy, Minakata, Katsumi, . . . Johnson, Walter W. (2011). Conflict resolution automation and pilot situation awareness *Human Interface and the Management of Information. Interacting with Information* (pp. 473-482): Springer.
- DiMeo, Karen, Kopardekar, Parimal, Ashford, Rose, Lozito, Sandy, & Mackintosh, Margaret-Anne. (2001). *Shared-separation: empirical results and theoretical implications*. Paper presented at the The Fourth International Air Traffic Management R&D Seminar ATM-2001.

- Dixon, Stephen R, & Wickens, Christopher D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(3), 474-486.
- Durso, Francis T, Bleckley, M Kathryn, & Dattel, Andrew R. (2006). Does situation awareness add to the validity of cognitive tests? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(4), 721-733.
- Durso, Francis T, & Manning, Carol A. (2008). Air traffic control. *Reviews of human factors and ergonomics*, 4(1), 195-244.
- Durso, Frank T, Dattel, Andrew R, Banbury, S, & Tremblay, Sebastien. (2004). SPAM: The real-time assessment of SA. In Simon Banburry & Sébastien Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (Vol. 1, pp. 137-154).
- Dwyer, John P, & Landry, Steven. (2009). Separation assurance and collision avoidance concepts for the next generation air transportation system *Human Interface and the Management of Information. Information and Interaction* (pp. 748-757): Springer.
- Dzindolet, Mary T, Peterson, Scott A, Pomranky, Regina A, Pierce, Linda G, & Beck, Hall P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6), 697-718.
- Endsley, Mica R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 32-64.
- Endsley, Mica R, & Garland, DJ. (2000). Theoretical underpinnings of situation awareness: A critical review. *Situation awareness analysis and measurement*, 3-32.
- Endsley, Mica R, & Kaber, D B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.
- Endsley, Mica R, Sollenberger, Randy, & Stein, Earl. (1999). *The use of predictive displays for aiding controller situation awareness*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

- Endsley, MR. (1997). *Communication and situation awareness in the aviation system*. Paper presented at the Aviation Communication: A Multi-Cultural Forum, Prescott, AZ.
- Erzberger, Heinz. (2006). *Automated conflict resolution for air traffic control*: [National Aeronautics and Space Administration], Ames Research Center.
- Erzberger, Heinz, & Heere, Karen. (2010). Algorithm and operational concept for resolving short-range conflicts. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 224(2), 225-243.
- Erzberger, Heinz, Paielli, Russell A, Isaacson, Douglas R, & Eshow, Michelle M. (1997). *Conflict detection and resolution in the presence of prediction error*. Paper presented at the 1st USA/Europe Air Traffic Management R&D Seminar, Saclay, France.
- EUROCONTROL. (2004). EUROCONTROL Operational Concept Document (OCD) (Vol. 1): European Organization for the Safety of Air Navigation.
- Eyferth, Klaus, Niessen, Cornelia, & Spaeth, Oliver. (2003). A model of air traffic controllers' conflict detection and conflict resolution. *Aerospace science and technology*, 7(6), 409-416.
- Fothergill, Selina, Loft, Shayne, & Neal, Andrew. (2009). ATC-labAdvanced: An air traffic control simulator with realism and control. *Behavior Research Methods*, 41(1), 118-127.
- Galster, Scott M, Duley, Jacqueline A, Masalonis, Anthony J, & Parasuraman, Raja. (2001). Air traffic controller performance and workload under mature free flight: Conflict detection and resolution of aircraft self-separation. *The International Journal of Aviation Psychology*, 11(1), 71-93.
- Gawron, Valerie J. (2008). *Human performance, workload, and situational awareness measures handbook*: CRC Press.
- Gibson, Jason, Orasanu, Judith, Villeda, Eric, & Nygren, Thomas E. (1997). *Loss of situation awareness- Causes and consequences*. Paper presented at the International Symposium on Aviation Psychology, 9 th, Columbus, OH.

- Goillau, P, Kelly, C, Boardman, M, & Jeannot, E. (2003). Guidelines for trust in future ATM systems: Measures: Document HRS/HSP-005-GUI-02. Brussels, Belgium. European Organisation for the Safety of Air Navigation.
- Goritzlehner, Reinier, Borst, Christopher, Ellerbroek, Joost, Westin, Carl, van Paassen, Marinus M, & Mulder, Max. (2014). *Effects of transparency on the acceptance of automated resolution advisories*. Paper presented at the Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on.
- Hancock, Peter A. (1999). *Human Performance and Ergonomics: Perceptual and Cognitive Principles*: Academic Press.
- Hart, Sandra G, & Staveland, Lowell E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, 1(3), 139-183.
- Heylen, FM, Van Dam, SBJ, Mulder, M, & van Paassen, MM. (2008). Design and evaluation of a vertical separation assistance display. *interface*, 19(22), 17.
- Higham, Tiana-Lynn M, Vu, Kim-Phuong L, Miles, J, Strybel, Thomas Z, & Battiste, Vernol. (2013). Training air traffic controller trust in automation within a NextGen environment. In S Yamamoto (Ed.), *HIMI/HCII 2013* (Vol. 8017, pp. 76-84). Heidelberg: Springer.
- Hilburn, Brian, & Jorna, P. G. A. M. (2001). Workload and Air Traffic Control. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 384-394). Hillsdale, New Jersey: Lawrence Erlbaum.
- Hilburn, Brian, Parasuraman, Raja, Jha, Pratik, & McGarry, Kathleen. (2006). Emerging Human Factors Issues in Future Air Traffic Management.
- Histon, JM, Hansman, RJ, Gottlieb, B, Kleinwaks, H, Yenson, S, Delahaye, D, & Puechmorel, S. (2002). *Structural considerations and cognitive complexity in air traffic control*. Paper presented at the Digital Avionics Systems Conference, 2002. Proceedings. The 21st.
- Hoc, Jean-Michel. (2001). Towards a cognitive approach to human-machine cooperation in dynamic situations. *International Journal of Human-Computer Studies*, 54(4), 509-540.

- Hoekstra, Jacco M, Van Gent, RNHW, & Ruigrok, RCJ. (2002). Designing for safety: the 'free flight' air traffic management concept. *Reliability Engineering & System Safety*, 75(2), 215-232.
- Hoff, Kevin Anthony, & Bashir, Masooda. (2015). Trust in Automation Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3), 407-434.
- Hoffman, Robert, Mukherjee, Avijit, & Vossen, Thomas WM. (2011). Air traffic management. *Wiley Encyclopedia of Operations Research and Management Science*.
- Hopkin, V David. (1991). The impact of automation on air traffic control systems *Automation and systems issues in air traffic control* (pp. 3-19): Springer.
- Hunt, Sarah Marie. (2013). The Impact of Trajectory Prediction Uncertainty on Reliance Strategy and Trust Attitude in an Automated Air Traffic Management Environment.
- ICAO. (2006). Report of the Second Meeting of Automatic Dependent Surveillance - Broadcast (ADS-B) Study and Implementation Task Force (ADS-B SI TF/2). Bangkok, Thailand: International Civil Aviation Organization.
- ICAO. (2010). Regional Report: Intensifying Asia-Pacific Collaboration to Address Efficiency and Safety: International Civil Aviation Organization Asia Pacific Office.
- Imbert, Jean-Paul, Hodgetts, Helen M, Parise, Robert, Vachon, François, Dehais, Frédéric, & Tremblay, Sébastien. (2014). Attentional costs and failures in air traffic control notifications. *Ergonomics*(ahead-of-print), 1-16.
- Isaacson, Douglas R, & Robinson III, John E. (2001). *A knowledge-based conflict resolution algorithm for terminal area air traffic control advisory generation*. Paper presented at the AIAA Guidance, Navigation, and Control Conference, Montreal, Canada.
- Ito, PK. (1980). 7 Robustness of ANOVA and MANOVA test procedures. *Handbook of statistics, 1*, 199-236.
- Jha, Pratik D, Bisantz, Ann M, Parasuraman, Raja, & Drury, Colin G. (2011). Air Traffic Controllers' Performance in Advance Air Traffic Management System: Part I—

- Performance Results. *The International Journal of Aviation Psychology*, 21(3), 283-305.
- Jian, Jiun-Yin, Bisantz, Ann M, & Drury, Colin G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1), 53-71.
- John, Mark St, & Manes, Daniel I. (2002). *Making unreliable automation useful*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Kallus, KW, Van Damme, D, & Dittman, A. (1999). Integrated job and task analysis of air traffic controllers: Phase 2. *Task analysis of en-route controllers (European Air Traffic Management Programme Rep. No. HUM. ET1. ST01. 1000-REP-04)*. EUROCONTROL, Brussels.
- Kaniasarasu, P, Steinfeld, A., Desai, M., & Yanco, H. (2012). *Potential measures for detecting trust changes*. Paper presented at the 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI).
- Kantowitz, Barry H, & Sorkin, Robert D. (1983). *Human factors: Understanding people-system relationships*. New York: John Wiley & Sons Inc.
- Kelly, C, Boardman, M, Goillau, P, & Jeannot, E. (2003). Guidelines for trust in future atm systems: A literature review *European Air Traffic Management Programme* (Vol. 1.0): European Organisation for the Safety of Air Navigation.
- Kennedy, Joshua S. (2005). A Human-Automation Interface Model to Guide Automation Design of System Functions: A Way to Achieve Manning Goals in New Systems: DTIC Document.
- Kontogiannis, Tom, & Malakis, Stathis. (2013). Strategies in controlling, coordinating and adapting performance in air traffic control: modelling 'loss of control' events. *Cognition, technology & work*, 15(2), 153-169.
- Kopardekar, P, & Magyarits, S. (2002). *Dynamic density: measuring and predicting sector complexity [ATC]*. Paper presented at the Digital Avionics Systems Conference, 2002. Proceedings. The 21st.

- Kuchar, James K, & Yang, Lee C. (2000). A review of conflict detection and resolution modeling methods. *Intelligent Transportation Systems, IEEE Transactions on*, 1(4), 179-189.
- Langan-Fox, Janice, Sankey, Michael J, & Canty, James M. (2009). Human factors measurement for future air traffic control systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 51(5), 595-637.
- Lee, John D. (2008). Review of a Pivotal Human Factors Article: "Humans and Automation: Use, Misuse, Disuse, Abuse". *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3), 404-410.
- Lee, John D, & Moray, Neville. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243-1270.
- Lee, John D, & See, Kathreen A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50-80. doi: 10.1518/hfes.46.1.50_30392
- Lee, John D, & Seppelt, Bobbie D. (2009). Human factors in automation design *Springer handbook of automation* (pp. 417-436): Springer.
- Li, Huiyang, Wickens, Christopher D, Sarter, Nadine, & Sebok, Angelia. (2014). Stages and Levels of Automation in Support of Space Teleoperations. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720814522830.
- Loft, Shayne, Bolland, S, & Humphreys, M. (2007). *Modeling the human air traffic controller, Part I: Expert-trainee differences in conflict detection*. Paper presented at the 14th International Symposium on Aviation Psychology.
- Loft, Shayne, Bowden, Vanessa, Braithwaite, Janelle, Morrell, Daniel B, Huf, Samuel, & Durso, Francis T. (2015). Situation awareness measures for simulated submarine track management. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(2), 298-310.
- Loft, Shayne, Sanderson, Penelope, Neal, Andrew, & Mooij, Martijn. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(3), 376-399.

- Madsen, Maria, & Gregor, Shirley. (2000). *Measuring human-computer trust*. Paper presented at the Proceedings of Eleventh Australasian Conference on Information Systems.
- Manzey, D, Reichenbach, J, & Onnasch, L. (2012). Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience. *Journal of Cognitive Engineering and Decision Making*, 6(1), 57-87. doi: 10.1177/1555343411433844
- Manzey, Dietrich, Gérard, Nina, & Wiczorek, Rebecca. (2014). Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload. *Ergonomics*(ahead-of-print), 1-23.
- Martin, C., & Imbert, Jean-Paul. (2012). *Introduction of a more automated environment in en-route air traffic control*. Paper presented at the SESAE Innovation Days.
- Marusich, Laura R, Bakdash, Jonathan Z, Onal, Emrah, Michael, S Yu, Schaffer, James, O'Donovan, John, . . . Gonzalez, Cleotilde. (2016). Effects of Information Availability on Command-and-Control Decision Making Performance, Trust, and Situation Awareness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58(2), 301-321.
- Masalonis, Anthony J, & Parasuraman, Raja. (1999). *Trust as a construct for evaluation of automated aids: Past and future theory and research*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Masalonis, Anthony J, & Parasuraman, Raja. (2003). *Effects of situation-specific reliability on trust and usage of automated air traffic control decision aids*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Maxwell, Scott E, & Delaney, Harold D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (Vol. 1): Psychology Press.
- McElreath, Kenneth W, & Pingsterhaus, Scott A. (2000). Integrated vertical situation display for aircraft: Google Patents.

- Mercer, Joey, Gomez, Ashley, Homola, Jeffrey, & Prevot, Thomas. (2014). *A closer look at automation behavior during a human-in-the-loop simulation*. Paper presented at the Digital Avionics Systems Conference (DASC), 2014 IEEE/AIAA 33rd.
- Merritt, Stephanie M, Lee, Deborah, Unnerstall, Jennifer L, & Huber, Kelli. (2014). Are Well-Calibrated Users Effective Users? Associations Between Calibration of Trust and Performance on an Automation-Aided Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720814561675.
- Merritt, Stephanie M., & Ilgen, Daniel R. (2008). Not All Trust Is Created Equal: Dispositional and History-Based Trust in Human-Automation Interactions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(2), 194-210. doi: 10.1518/001872008x288574
- Merwin, David H, & Wickens, Christopher D. (1996). Evaluation of Perspective and Coplanar Cockpit Display of Traffic Information to Support Hazard Awareness in Free Flight *University of Illinois Institute of Aviation Technical Report ARL-96-5/NASA-96-1*: Aviation Research Laboratory, Savoy, Illinois.
- Metzger, Ulla, & Parasuraman, Raja. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(4), 519-528.
- Metzger, Ulla, & Parasuraman, Raja. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(1), 35-49.
- Metzger, Ulla, & Parasuraman, Raja. (2006). Effects of automated conflict cuing and traffic density on air traffic controller performance and visual attention in a datalink environment. *The International Journal of Aviation Psychology*, 16(4), 343-362.
- Meyer, Joachim. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(4), 563-572.

- Meyer, Joachim. (2004). Conceptual issues in the study of dynamic hazard warnings. *HUMAN FACTORS-NEW YORK THEN SANTA MONICA*-, 46, 196-204.
- Mogford, Richard H, Guttman, JA, Morrow, SL, & Kopardekar, P. (1995). The Complexity Construct in Air Traffic Control: A Review and Synthesis of the Literature: DTIC Document.
- Molloy, Robert, & Parasuraman, Raja. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38(2), 311-322.
- Montgomery, Douglas C. (2008). *Design and analysis of experiments*: John Wiley & Sons.
- Mosier, Kathleen L, & Fischer, Ute M. (2010). Judgment and decision making by individuals and teams: Issues, models, and applications. *Reviews of human factors and ergonomics*, 6(1), 198-256.
- Mosier, Kathleen L, & Skitka, Linda J. (1996). Human decision makers and automated decision aids: Made for each other. *Automation and human performance: Theory and applications*, 201-220.
- Mosier, Kathleen L, Skitka, Linda J, Heers, Susan, & Burdick, Mark. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International journal of aviation psychology*, 8(1), 47-63.
- Murphy, Elizabeth D, Albert, Harold A, Chen, Jennifer M, & Anderson, Gregory G. (2012). *The Role of Mental Computations in Current and Future En Route Air Traffic Control*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Nass, Clifford, Fogg, BJ, & Moon, Youngme. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, 45(6), 669-678.
- Neal, Andrew, Hannah, Sam, Sanderson, Penelope, Bolland, Scott, Mooij, Martijn, & Murphy, Sean. (2013). Development and validation of a multilevel model for predicting workload under routine and nonroutine conditions in an air traffic management center. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720813491283.

- Nguyen, Jimmy H, Bacon, L Paige, Rorie, R Conrad, Herron, Meghann, Vu, Kim-Phuong L, Strybel, Thomas Z, & Battiste, Vernol. (2011). How data comm methods and multi-dimensional traffic displays influence pilot workload under trajectory based operations *Human Interface and the Management of Information. Interacting with Information* (pp. 507-515): Springer.
- Niessen, Cornelia, Eyferth, Klaus, & Bierwagen, T. (1999). Modelling cognitive processes of experienced air traffic controllers. *Ergonomics*, 42(11), 1507-1520.
- Nijhuis, H. (1999). WP8: Summary and consolidation of RHEA results *report RHEA/NL/WPR/8/04*: European Commission DG VII.
- Nolan, Michael S. (1990). *Fundamentals of air traffic control*. Belmont: Wadsworth, Inc.
- Norman, Donald A. (1991). Approaches to the study of intelligence. *Artificial Intelligence*, 47(1), 327-346.
- Nunes, Ashley. (2003). *The impact of automation use on the mental model: findings from the air traffic control domain*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Nunes, Ashley, & Mogford, Richard H. (2003). *Identifying controller strategies that support the 'picture'*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Olmos, Oscar, Wickens, Christopher D, & Chudy, Andrew. (2000). Tactical displays for combat awareness: An examination of dimensionality and frame of reference concepts and the application of cognitive engineering. *The International Journal of Aviation Psychology*, 10(3), 247-271.
- Onnasch, Linda, Ruff, Stefan, & Manzey, Dietrich. (2014). Operators' adaptation to imperfect automation—impact of miss-prone alarm Systems on attention allocation and performance. *International Journal of Human-Computer Studies*.
- Onnasch, Linda, Wickens, Christopher D, Li, Huiyang, & Manzey, Dietrich. (2014). Human Performance Consequences of Stages and Levels of Automation An Integrated Meta-Analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(3), 476-488.

- Parasuraman, Raja, & Hancock, Peter A. (2001). Adaptive Control of Mental Workload. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 305-320). New Jersey: Erlbaum.
- Parasuraman, Raja, & Manzey, Dietrich H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(3), 381-410.
- Parasuraman, Raja, & Miller, C.A. (2004). Trust and etiquette in high-criticality automated systems. *Communications of the ACM*, 47(4), 51-55.
- Parasuraman, Raja, Molloy, Robert, & Singh, Indramani L. (1993). Performance consequences of automation-induced 'complacency'. *The International Journal of Aviation Psychology*, 3(1), 1-23.
- Parasuraman, Raja, & Riley, Victor. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, 39(2), 230-253.
- Parasuraman, Raja, Sheridan, Thomas B, & Wickens, Christopher D. (2000). A Model for Types and Level of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 30(3), 286-289.
- Pawlak, William S, Brinton, Christopher R, Crouch, Kimberly, & Lancaster, Kenneth M. (1996). *A framework for the evaluation of air traffic control complexity*. Paper presented at the Proceedings of the AIAA Guidance Navigation and Control Conference, San Diego, CA.
- Perry, Tekla S. (1997). In search of the future of air traffic control. *Spectrum, IEEE*, 34(8), 18-35.
- Prevot, Thomas. (1998). *A Display for Managing the Vertical Flight Path—an Appropriate Task with Inappropriate Feedback*. Paper presented at the Proc. HCI Aero'98.
- Prevot, Thomas, Homola, J. R., Martin, L. H., Mercer, J. S., & Cabrall, C. D. (2012). Toward Automated Air Traffic Control—Investigating a Fundamental Paradigm Shift in Human/Systems Interaction. *International Journal of Human-Computer Interaction*, 28(2), 77-98.
- Prevot, Thomas, Homola, Jeffrey, & Mercer, Joey. (2008). Human-in-the-loop evaluation of ground-based automated separation assurance for NEXTGEN. *Proceedings of*

- the 26th International Congress of the Aeronautical Sciences*, AIAA-ATIO-2008-8885.
- Prevot, Thomas, Mercer, J., Martin, L., Homola, J., Cabrall, C., & Brasil, C. (2010). *Evaluation of NextGen Air Traffic Control Operations with Automated Separation Assurance*. Paper presented at the International Conference on Human-Computer Interaction in Aerospace, Cape Canaveral, Florida.
- Prevot, Thomas, & Palmer, EA. (2000). *Staying ahead of the automation: A vertical situation display can help*. Paper presented at the World Aviation Conference, San Diego, CA.
- Pritchett, Amy R. (2009). Aviation automation: General perspectives and specific guidance for the design of modes and alerts. *Reviews of human factors and ergonomics*, 5(1), 82-113.
- Proud, Ryan W, Hart, Jeremy J, & Mrozinski, Richard B. (2003). Methods for determining the level of autonomy to design into a human spaceflight vehicle: a function specific approach: DTIC Document.
- Rantanen, Esa M. (2004). Development and validation of objective performance and workload measures in air traffic control (Aviation Human Factors Division AHFD, Trans.). Illinois: Institute of Aviation, University of Illinois at Urbana-Champaign.
- Rantanen, Esa M, & Nunes, Ashley. (2005). Hierarchical conflict detection in air traffic control. *The International Journal of Aviation Psychology*, 15(4), 339-362.
- Rantanen, Esa M, & Wickens, Christopher D. (2012). Conflict Resolution Maneuvers in Air Traffic Control: Investigation of Operational Data. *The International Journal of Aviation Psychology*, 22(3), 266-281.
- Rehmann, Albert J. (1995). Handbook of Human Performance Measures and Crew Requirements for Flightdeck Research: DTIC Document.
- Rice, Stephen. (2009). Examining Single and Multiple Process Theories of Trust in Automation. *The Journal of General Psychology*, 136(3), 303.
- Rodgers, Mark D, Mogford, Richard H, & Strauch, Barry. (2000). Post hoc assessment of situation awareness in air traffic control incidents and major aircraft accidents. *Situation awareness analysis and measurement*, 73-112.

- Rong, Jie, Geng, Shijian, Valasek, John, & Ioerger, Thomas R. (2002). *Air traffic conflict negotiation and resolution using an onboard multi-agent system*. Paper presented at the Digital Avionics Systems Conference, 2002. Proceedings. The 21st.
- Rovira, Ericka, Cross, Austin, Leitch, Evan, & Bonaceto, Craig. (2014). Displaying Contextual Information Reduces the Costs of Imperfect Decision Automation in Rapid Retasking of ISR Assets. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720813519675.
- Rovira, Ericka, McGarry, Kathleen, & Parasuraman, Raja. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(1), 76-87.
- Rovira, Ericka, & Parasuraman, Raja. (2010). Transitioning to future air traffic management: Effects of imperfect automation on controller attention and performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(3), 411-425.
- Salmon, Paul, Stanton, Neville, Walker, Guy, & Green, Damian. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied ergonomics*, 37(2), 225-238.
- Sanchez, Julian, Krumbholz, Cole J, & Silva, Marcio O. (2009). Controller-Automation Interaction in NextGen: A New Paradigm: The MITRE Corporation.
- Sanchez, Julian, Rogers, Wendy A, Fisk, Arthur D, & Rovira, Ericka. (2014). Understanding reliance on automation: effects of error type, error distribution, age and experience. *Theoretical Issues in Ergonomics Science*, 15(2), 134-160.
- Sarter, Nadine. (2008). Investigating mode errors on automated flight decks: Illustrating the problem-driven, cumulative, and interdisciplinary nature of human factors research. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3), 506-510.
- Schaefer, Kristin E, Chen, Jessie YC, Szalma, James L, & Hancock, PA. (2016). A Meta-Analysis of Factors Influencing the Development of Trust in Automation Implications for Understanding Autonomy in Future Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720816634228.

- Sebok, Angelia, & Wickens, C. D. (2016). Implementing lumberjacks and black swans into model-based tools to support human-automation interaction. *Human factors*, (ahead-of-print).
- SESAR. (2015). SJU Consolidated Annual Activity Report. In Florian Guillermet (Ed.).
- Sheridan, Thomas B. (2002). *Humans and automation: System design and research issues*: John Wiley & Sons, Inc.
- Sheridan, Thomas B. (2006). *Next Generation Air Transportation Systems: Human-Automation Interaction and Organizational Risks*. Paper presented at the Proceedings of the 2nd Symposium on Resilience Engineering.
- Sheridan, Thomas B, & Parasuraman, Raja. (2005). Human-automation interaction. *Reviews of human factors and ergonomics*, 1(1), 89-129.
- Shneiderman, Shneiderman Ben, & Plaisant, Catherine. (2005). Designing the user interface 4 th edition. ed: *Pearson Addison Wesley, USA*.
- Singh, Indramani L, Molloy, Robert, & Parasuraman, Raja. (1997). Automation-induced monitoring inefficiency: role of display location. *International Journal of Human-Computer Studies*, 46(1), 17-30.
- Sridhar, Banavar, Sheth, Kapil S, & Grabbe, Shon. (1998). *Airspace complexity and its application in air traffic management*. Paper presented at the 2nd USA/Europe Air Traffic Management R&D Seminar.
- Stankovic, Stéphanie, Raufaste, Éric, & Averty, Philippe. (2008). Determinants of conflict detection: A model of risk judgments in air traffic control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(1), 121-134.
- Suijkerbuijk, HCH, Borst, C, Mulder, M, & Van Paassen, MM. (2005). *Development and Experimental Evaluation of a Performance-Based Vertical Situation Display*. Paper presented at the 2005 AIAA Guidance, Navigation, and Control Conference and Exhibit.
- Tham, M., & Wickens, C. D. (1993). Evaluation of Perspective and Stereoscopic Displays as Alternatives to Plan View Displays in Air Traffic Control. University of Illinois Institute of Aviation Technical Report (ARL-93-4/FAA-93-1): Aviation Research Laboratory, Savoy, Illinois.

- Thomas, Lisa C, & Rantanen, Esa M. (2006). Human factors issues in implementation of advanced aviation technologies: A case of false alerts and cockpit displays of traffic information. *Theoretical Issues in Ergonomics Science*, 7(5), 501-523.
- Thomas, Lisa C, & Wickens, C. D. (2005). Effects of Display Dimensionality, Conflict Geometry, and Time Pressure on Conflict Detection and Resolution Performance Using a Cockpit Display of Traffic Information (Aviation Human Factors Division, Trans.). Savoy, Illinois: Institute of Aviation, University of Illinois at Urbana-Champaign.
- Thomas, Lisa C, & Wickens, Christopher D. (2006). Display dimensionality, conflict geometry, and time pressure effects on conflict detection and resolution performance using cockpit displays of traffic information. *The International Journal of Aviation Psychology*, 16(3), 321-342.
- Thomas, Lisa C, Wickens, Christopher D, & Rantanen, Esa M. (2003). *Imperfect automation in aviation traffic alerts: A review of conflict detection algorithms and their implications for human factors research*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Trapsilawati, Fitri, Qu, Xingda, Wickens, Chris D, & Chen, Chun-Hsien. (2015). Human factors assessment of conflict resolution aid reliability and time pressure in future air traffic control. *Ergonomics*, 58(6), 897-908.
- USDOT. (2014). Number of Pilot-Reported Near Midair Collisions (NMAC) by Degree of Hazard. from US Department of Transportation http://www.rita.dot.gov/bts/sites/rita.dot.gov/bts/files/publications/national_transportation_statistics/html/table_02_15.html
- Vakil, Sanjay S, Midkiff, Alan H, & Hansman, R John. (1996). Development and evaluation of an electronic vertical situation display: International Center for Air Transportation.
- Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2012). Trust in Smart Systems: Sharing Driving Goals and Giving Information to Increase Trustworthiness and Acceptability of Smart Systems in Cars. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(5), 799-810. doi: 10.1177/0018720812443825

- Vossen, Thomas WM, Hoffman, Robert, & Mukherjee, Avijit. (2012). Air traffic flow management *Quantitative problem solving methods in the airline industry* (pp. 385-453): Springer.
- Vu, Kim-Phuong L, Strybel, Thomas Z, Battiste, Vernol, Lachter, Joel, Dao, Arik-Quang V, Brandt, Summer, . . . Johnson, Walter. (2012). Pilot performance in trajectory-based operations under concepts of operation that vary separation responsibility across pilots, air traffic controllers, and automation. *International Journal of Human-Computer Interaction*, 28(2), 107-118.
- Wang, Lu, Jamieson, Greg A. , & Hollands, Justin G. . (2009). Trust and Reliance on an Automated Combat Identification System. *Human Factors*, 51(3), 281-291. doi: 10.1177/0018720809338842
- Warm, Joel S, Parasuraman, Raja, & Matthews, Gerald. (2008). Vigilance requires hard mental work and is stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3), 433-441.
- Weinstock, A, Oron-Gilad, T, & Parmet, Y. (2012). The effect of system aesthetics on trust, cooperation, satisfaction and annoyance in an imperfect automated system. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 41, 258-265.
- Westin, C A, Borst, C., & Hilburn, B. (2013). *Mismatches between automation and human strategies: an investigation into future air traffic management decision aiding*. Paper presented at the International Symposium on Aviation Psychology, Ohio, USA.
- Wickens, C. D., Mavor, AS, Parasuraman, R, & McGee, P. (1998). Airspace system integration: The concept of free flight. *The future of air traffic control*.
- Wickens, Christopher D. (2000a). The trade-off of design for routine and unexpected performance: Implications of situation awareness. In Daniel J Garland & Mica Endsley (Eds.), *Situation awareness analysis and measurement*. (pp. 211-225). Mahwah, New Jersey: Lawrence Erlbaum.
- Wickens, Christopher D. (2000b). *The when and how of using 2-D and 3-D displays for operational tasks*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

- Wickens, Christopher D, Alexander, Amy L, Ambinder, Michael S, & Martens, Marieke. (2004). The role of highlighting in visual search through maps. *Spatial vision*, 17(4), 373-388.
- Wickens, Christopher D, Campbell, M, Liang, CC, & Merwin, DH. (1995). Weather displays for air traffic control: The effect of 3D perspective. *University of Illinois Institute of Aviation Technical Report (ARL-95-1/FAA-95-1)*. Savoy, IL: Aviation Res. Lab.
- Wickens, Christopher D, Clegg, Benjamin A, Vieane, Alex Z, & Sebok, Angelia L. (2015). Complacency and automation bias in the use of imperfect automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(5), 728-739.
- Wickens, Christopher D, Conejo, Rena, & Gempier, Keith. (1999). *Unreliable automated attention cueing for air-ground targeting and traffic maneuvering*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Wickens, Christopher D, & Dixon, Stephen R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3), 201-212.
- Wickens, Christopher D, Gempier, Keith, & Morpew, M Ephimia. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors*, 2(2), 99-126.
- Wickens, Christopher D, Hollands, J.G., Banbury, S., & Parasuraman, R. (2013) *Engineering Psychology and Human Performance* (4 ed.). New Jersey: Pearson Education, Inc.
- Wickens, Christopher D, Hollands, Justin G, Banburry, Simon, & Parasuraman, Raja. (2013). *Engineering Psychology and Human Performance* (fourth ed.). New Jersey: Pearson Education Inc.
- Wickens, Christopher D, Kroft, Paul, & Yeh, Michelle. (2000). *Data base overlay in electronic map design: Testing a computational model*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

- Wickens, Christopher D, Mavor, Anne S, Parasuraman, R, & McGee, P. (1998). *The future of air traffic control: Human operators and automation*. Washington D.C: National Academies Press.
- Wickens, Christopher D, & May, Patricia. (1994). Terrain representation for air traffic control: A comparison of perspective with plan view displays. *University of Illinois Institute of Aviation Technical Report ARL-94-10/FAA-94-2*. Savoy, IL.
- Wickens, Christopher D, Miller, S., & Tham, M. (1996). The implications of data-link for representing pilot request information on 2d and 3d air traffic control displays. *International Journal of Industrial Ergonomics*, 18(4), 283-293.
- Wickens, Christopher D, Rice, Stephen, Keller, David, Hutchins, Shaun, Hughes, Jamie, & Clayton, Krisstal. (2009). False Alerts in Air Traffic Control Conflict Alerting System: Is There a “Cry Wolf” Effect? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 51(4), 446-462.
- Wickens, Christopher D, & Xu, Xidong. (2002). *Automation trust, reliability and attention HMI 02-03*: University of Illinois at Urbana-Champaign, Aviation Human Factors Division.
- Wiener, Earl L, & Curry, Renwick E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23(10), 995-1011.
- Willems, B, & Truitt, T. R. (1999). Implications of reduced involvement in en route air traffic control: DTIC Document.
- Willems, Ben, & Heiney, M. (2002). Decision support automation research in the en route air traffic control environment: DTIC Document.
- Wise, John A, Hopkin, V David, & Garland, Daniel J. (2010). *Handbook of aviation human factors*: CRC PressI Llc.
- Woods, David D. (1984). Visual momentum: a concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies*, 21(3), 229-244.
- Xing, Jing, & Manning, Carol A. (2005). Complexity and automation displays of air traffic control: Literature review and analysis: DTIC Document.

- Xu, Xidong. (2003). *Conflict Detection with Cockpit Display of Traffic Information: What is it, What has Been Found, and What Needs to be Done?* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Yeh, Michelle, Merlo, James L, Wickens, Christopher D, & Brandenburg, David L. (2003). Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(3), 390-407.
- Yin, Shanqing, Wickens, C. D., Helander, M., & Laberge, J. C. (2015). Predictive Displays for a Process-Control Schematic Interface. *Human factors*, 57, 110-124.
- Young, Mark S, Brookhuis, Karel A, Wickens, Christopher D, & Hancock, Peter A. (2014). State of science: mental workload in ergonomics. *Ergonomics*, 1(58), 1-17.

Appendixes

Appendix A. List of maneuvering abbreviations and the units

Notation	Abbreviation	Unit
C	Climb	Flight level
D	Descend	Flight level
FH	Fly heading	Degree
IS	Increase speed	Knots
RS	Reduce speed	Knots

Appendix B. List of probe questions

Level 1

- 1 What is the aircraft A's speed?
- 2 What is the direction of departure for aircraft B?
- 3 What is the altitude clearance for aircraft C?

Level 2

- 1 How many aircraft flying southbound?
- 2 Which aircraft has lower altitude?
- 3 Is the difference in heading between aircraft D and aircraft E more than 90°?

Level 3

- 1 Which aircraft must be handed-off to another sector within the next 2 minutes?
- 2 Which pairs of aircraft will lose separation if they stay on their current courses?
- 3 Which aircraft will need a new clearance to achieve landing requirements?

Appendix C. an Example of NASA-TLX Application (Hart and Staveland, 1998)

EXAMPLE:

COMPARE WORKLOAD OF TWO TASKS THAT REQUIRE A SERIES OF DISCRETE RESPONSES. THE PRIMARY DIFFICULTY MANIPULATION IS THE INTER-STIMULUS INTERVAL (ISI) - (TASK 1 = 500 msec. TASK 2 = 300 msec)

PAIR-WISE COMPARISONS OF FACTORS:

INSTRUCTIONS: SELECT THE MEMBER OF EACH PAIR THAT PROVIDED THE MOST SIGNIFICANT SOURCE OF WORKLOAD VARIATION IN THESE TASKS

TALLY OF IMPORTANCE SELECTIONS		
PD / MD	TD / PD	TD / FR
TD / MD	OP / PD	TD / EF
OP / MD	FR / PD	OP / FR
FR / MD	EF / PD	OP / EF
EF / MD	TD / OP	EF / FR
MD III = 3 PD = 0 TD IIII = 5 OP I = 1 FR III = 3 EF III = 3 SUM = 15		

RATING SCALES:

INSTRUCTIONS: PLACE A MARK ON EACH SCALE THAT REPRESENTS THE MAGNITUDE OF EACH FACTOR IN THE TASK YOU JUST PERFORMED

DEMANDS	RATINGS FOR TASK 1:		RATING	WEIGHT	PRODUCT
MD	LOW	I <u> x </u> I	HIGH	30	x 3 = 90
PD	LOW	I <u> x </u> I	HIGH	15	x 0 = 0
TD	LOW	I <u> x </u> I	HIGH	60	x 5 = 150
OP	EXCL	I <u> x </u> I	POOR	40	x 1 = 40
FR	LOW	I <u> x </u> I	HIGH	30	x 3 = 90
EF	LOW	I <u> x </u> I	HIGH	40	x 3 = 120
SUM					= 490
WEIGHTS (TOTAL)					= 15
MEAN WWL SCORE					= 32

DEMANDS	RATINGS FOR TASK 2:		RATING	WEIGHT	PRODUCT
MD	LOW	I <u> x </u> I	HIGH	30	x 3 = 90
PD	LOW	I <u> x </u> I	HIGH	25	x 0 = 0
TD	LOW	I <u> x </u> I	HIGH	70	x 5 = 350
OP	EXCL	I <u> x </u> I	POOR	50	x 1 = 50
FR	LOW	I <u> x </u> I	HIGH	50	x 3 = 150
EF	LOW	I <u> x </u> I	HIGH	30	x 3 = 90
SUM					= 730
WEIGHTS (TOTAL)					= 15
MEAN WWL SCORE					= 49

RESULTS:

SUBSCALES PINPOINT SPECIFIC SOURCE OF WORKLOAD VARIATION BETWEEN TASKS (TD). THE WWL SCORE REFLECTS THE IMPORTANCE OF THIS AND OTHER FACTORS AS WORKLOAD-DRIVERS AND THEIR SUBJECTIVE MAGNITUDE IN EACH TASK

Appendix D. Trust Questionnaire (Jian, Bisantz and Drury, 2000)

Below is a list of statements for evaluating trust between people and automation. There are several scales for you to rate intensity of your feeling of trust, or your impression of the system while operating a machine. Please mark an "x" on each line at the point which best describes your feeling or your impression.

(Note: not at all=1; extremely=7)

- 1 The system is deceptive
 1 2 3 4 5 6 7
- 2 The system behaves in an underhanded manner
 1 2 3 4 5 6 7
- 3 I am suspicious of the system's intent, action, or outputs
 1 2 3 4 5 6 7
- 4 I am wary of the system
 1 2 3 4 5 6 7
- 5 The system's actions will have a harmful or injurious outcome
 1 2 3 4 5 6 7
- 6 I am confident in the system
 1 2 3 4 5 6 7
- 7 The system provides security
 1 2 3 4 5 6 7
- 8 The system has integrity
 1 2 3 4 5 6 7
- 9 The system is dependable
 1 2 3 4 5 6 7
- 10 The system is reliable
 1 2 3 4 5 6 7
- 11 I can trust the system
 1 2 3 4 5 6 7
- 12 I am familiar with the system
 1 2 3 4 5 6 7

Appendix E. List of SA Requirements for En Route ATC (Endsley and Rodgers, 1994)

SA Level	SA Requirements		
Level 1	Aircraft	- Received by correct aircraft	Airports
	- Aircraft ID, CID, beacon code	- Readback	- Operational status
	- Current route (position, heading, aircraft turn rate, altitude, climb/descent rate, groundspeed)	- correct/complete	- Restrictions in effect
	- Current flight plan (destination, filed plan)	- Pilot acceptance of clearance	- Direction of departures
	- Aircraft capabilities (turn rate, climb/descent rate, cruising speed, max/min speed)	- Flight progress strip current	- Current aircraft arrival rate
	- Equipment on board	Sector	- Arrival requirements
	- Aircraft type	- Special airspace status	- Active runways/approach
	- Fuel/loading	- Equipment functioning	- Sector saturation
	- Aircraft status	- Restrictions in effect	- Aircraft in holding (time, number, direction, leg length)
	Emergencies	- Changes to standard procedures	Weather
	- Type of emergency	Special Operations	- Area affected
	- Time on fuel remaining	- Type of special operation	- Altitudes affected
	- Souls on board	- Time begin/terminate operations	- Conditions (snow, icing, fog, hail, rain, turbulence, overhangs)
	Requests	- Projected duration	- Temperatures
	- Pilot/controller requests	- Area and altitude affected	- Intensity
	- Reason for request	ATC Equipment Malfunctions	- Visibility
	Clearances	- Equipment	- Winds
	- Assignment given	- position/range	- IFR/VFR conditions
		- Aircraft in outage area	- Airport conditions
Level 2	Conformance	Timing	Other
	- Amount of deviation (altitude, airspeed, route)	- Projected time in airspace	Sector/Airspace
	- Time until aircraft reaches assigned altitude, speed, route/heading	- Projected time till clear of airspace	- Impact of requests/clearances
	Current Separation	- Time until aircraft landing expected	- Impact of weather
	- Amount of separation between	- Time/distance aircraft to airport	Confidence
		- Order/sequencing of aircraft	Level/Accuracy of Info
			- Aircraft ID, position, altitude, airspeed, heading

	aircraft/objects/airspace/ground along route - Deviation between separation and prescribed limits - Number/timing aircraft on route - Altitudes available	Deviations - Deviation aircraft/landing requests - Deviation aircraft/flight plan - Deviation aircraft/pilot requests	- Weather - Altimeter setting
Level 3	Projected Aircraft Route (Current) - Position, flight plan, destination, heading, route, altitude, climb/descent rate, airspeed winds, groundspeed, intentions, assignments. Projected Aircraft Route (Potential) - Projected position x at time t - Potential assignments Projected Separation - Amount of separation along route (aircraft/objects/airspace/ground)	- Deviation between separation and prescribed limits - Relative projected aircraft routes - Relative timing along route Predicted Changes in Weather - Direction/speed of movement - Increasing/decreasing in intensity Impact of Potential Route Changes - Type of change required - Time and distance till turn aircraft	- Amount of turn/new heading, altitude route change required - Aircraft ability to make change - Projected no. of changes - Increase/decrease length of route - Cost/benefit of new clearance - Impact of proposed change

Appendix F. ATC Task Procedures

ATCSimulator®2

Arrival Procedures

Overview

Before aircraft can enter your area, a letter of agreement between facilities must exist which describe the procedures to allow entry. In ATCS2, the agreement exists that all aircraft will enter and continue on their route of flight to their destination airports. If you do not accept them by the time they reach their destination, they will maintain their current altitude and circle the airport.

Guiding the Arrival Aircraft to its destination

Visual or Instrument Approach?

Before directing an aircraft to its destination airport, you must determine the type of approach sequence that the aircraft can use. There are 2 types supported in ATCSimulator®2, Visual and/or Instrument. A visual approach allows the pilot to report to you when they can actually see the airport. And a visual approach is used on a runway where an instrument approach is not available. Please refer to the spoken phrases to issue an approach.

Clearing for a Visual Approach

Before clearing an aircraft for a visual approach, the aircraft will need to descend to an altitude that is at or below 5000 feet above ground level (AGL) before they can visually SEE the airport. When the aircraft is approximately 10 miles from the destination airport, he will report "I have the airport in sight". After the report has been received from the pilot, you will need to descend the aircraft to the proper approach altitude for a visual approach, which is at or below 3000 feet AGL. A good rule of thumb is to use 2500 to 3000 feet AGL when descending the target for the landing clearance. the elevation is for instance, 5352 feet MSL, a good altitude to use is $5352' + 2500' = 7852'$. You can then round down to the nearest hundredth of 7800 feet MSL. But, before assigning this altitude to the aircraft, you should consider the terrain in its path to see that there is nothing to interfere with that altitude.

Clearing for an Instrument Approach

The instrument approach is a pre-defined route which will allow the target to fly to a point that places the aircraft in line with the runway heading at a predetermined distance from

the landing threshold of the runway. This type of approach will ensure that all aircraft follow the same path, and allow the controller to use speed and altitude to affect separation between arriving aircraft. The instrument approach should be given at any time before the aircraft reaches the outer marker on the approach path.

If the aircraft is somewhere between the outer marker and the landing threshold of the runway, the aircraft will circle back around to intercept the course at a point 6 miles away from the outer marker. If you have allowed a target to come that close to the runway, issuing a visual approach would be appropriate for that target at that time. So, it is a good practice to issue the approach clearance somewhere beyond about 20 miles from the airport to allow for a smooth transition to intercept the approach path.

Initiating a Handoff to the Tower – Step 1 of 2

Whether it is a visual or instrument approach, prior to having the pilot of an aircraft contact the Tower, you must initiate a sequence, which makes the target visible on the Tower controller's radarscope. And this must be performed before issuing a frequency change to the pilot. Once the pilot has changed his radio frequency to the Towers, you can no longer communicate with the aircraft. So it is important that you do not issue a frequency change until all clearances have been delivered to the aircraft to affect a landing.

Handoff – Transfer of Communications – Step 2 of 2

Once the target has passed the outer marker and is lined up with the runway, issue a frequency change to the pilot. At that time, the datatag and history trail will vanish, and the lone target symbol will continue on its path until the aircraft has landed or dropped below 1000 feet AGL.

Departure Procedures

Overview

Each Departure has a "Proposed" time of Departure printed on its flight strip. This time is preceded with the letter "P" on the flight strip. The Tower will release the Departure automatically, instructing him to contact the Departure Controller (you) after takeoff. The system is designed to automatically display the full datablock once the target reaches a certain altitude. The Pilot will then contact you with a report of the altitude he is climbing to. It will then be up to you to direct the aircraft to its exit gate in your airspace, climb the target to the appropriate altitude, issue a frequency change and handoff to the Center controller. In most instances, you should not have to do anything except climb the aircraft.

Altitude Assignment

On initial contact by the Departure, the Pilot will report its assigned interim altitude. He will climb to that altitude and follow his filed flight plan to leave your airspace. If you will look on the associated aircrafts flight-strip, you will find the filed en-route altitude that the aircraft wishes to fly. Be sure you have issued the climb to this altitude before allowing the aircraft to leave your airspace. If you fail to issue an altitude change, the Pilot will continue at its present altitude when leaving your airspace.

Handoff to Center – Transfer of Communications

Once the Center controller has accepted the target, issue a frequency change to the pilot before he reaches his exit fix. To determine the correct frequency to issue to the pilot, review space 9 on the departure strip for this aircraft. After issuing the frequency change, the datatag will be reduced to a limited datablock (LDB) and will only display the destination and current altitude of the aircraft. Once reaching the exit fix, this target will disappear from your radarscope.

ATCSimulator®2 Command Phraseology

aaa = altitude in hundreds of feet (i.e. 8500=085, 12000=120, etc.)

sss = speed in knots

hhh = heading (i.e. 001 to 360)

*Note: always express using 3 digits

rrr = runway identifier (i.e. 35C = 3 5 center)

fff.f = frequency (i.e. 118.0)

*Note: always express as “point”, no whole numbers

mmm = miles

nnnnn = fix name (i.e. DOMAN intersection, ABQ vortac)

Spoken Phrase	Additional Phrase
Descend and maintain aaa	altitude
Climb and maintain aaa	altitude
Increase speed to sss	speed
Reduce speed to sss	speed
Fly heading hhh	degrees
Cleared Visual Approach Runway rrr	runway number
Cleared I L S Approach Runway rrr	runway number
Contact Center fff.f	frequency
Contact Tower fff.f	frequency
Turn Right heading hhh	degrees
Turn Left heading hhh	degrees
Proceed direct nnnnn	fix name
Cleared to nnnnn Hold as Published	fix name + HA
Say Speed	

Say Altitude	
Say Heading	
Report The Airport In Sight	
You are mmm miles from the nnnn	miles
Maintain aaa	altitude

NLR Air Traffic Simulator (NARSIM)

Data Block or Data Tag

Color Coding

Color Coding of Labels

- Salmon : Aircraft under control by another station/sector
- Red : Aircraft in a predicted conflict
- White : Aircraft under your control
- Pink : Aircraft in transfer to or from your station/sector

Color Coding of Strips

- Blue : Departures from Changi
- Grey : Overflights
- Yellow : Arrivals to Changi

Communication Sequence

Departures:

Singapore Approach → Singapore Approach North/South → Sector

Arrivals:

Sector → Singapore Approach North/South → Changi Tower

Departure Procedure

<u>APPROACH POSITION</u>	ATC Input
<p>AIRCRAFT: Singapore Approach, <CALLSIGN> climbing 7000 feet</p> <p>ATC: <CALLSIGN>, continue climb 7000 feet</p> <p>AIRCRAFT: <CALLSIGN> continuing climb 7000 feet. (when reached 8000 feet)</p> <p>ATC: <CALLSIGN> climb flight level 290 (or requested flight level), Contact Radar 123.7</p> <p>AIRCRAFT: <CALLSIGN> climbing flight level 290 (or requested flight level), contacting Radar 123.7</p>	<p>Click right hand level indicator and choose advised level</p> <p>Click call sign choose TRANSFER</p>

Arrival Procedure

<u>SECTOR POSITION</u>	ATC Input
<p>AIRCRAFT: Singapore Radar <CALLSIGN> maintaining <FLIGHTLEVEL>, <STANDARD ARRIVAL (NAME)> , Information "OSCAR" received</p> <p>ATC: <CALLSIGN>, Singapore Radar, identified, descend 10000 feet</p> <p>AIRCRAFT: <CALLSIGN>, descending 10000 feet (when approaching the 50nm ring)</p> <p>ATC: <CALLSIGN>, Singapore Radar, maintain 10000 feet contact Approach 124.4</p>	

<u>APPROACH POSITION</u>	ATC Input
<p>AIRCRAFT: Singapore Approach, <CALLSIGN> approaching 9000 feet request descent.</p> <p>ATC: <CALLSIGN> Singapore Approach, descend 6000 feet.</p> <p>AIRCRAFT: <CALLSIGN>, descending 6000 feet</p>	<p>Click right hand level indicator and choose advised level</p>

<u>ARRIVAL POSITION</u>	ATC Input
<p>AIRCRAFT: Singapore Arrival, <CALLSIGN> maintain 6000 feet</p> <p>ATC: <CALLSIGN> descend 3000 feet, radar vectors for Runway 20R</p> <p>AIRCRAFT: <CALLSIGN>, descend 3000 feet , radar vectors Runway 20R</p> <p>ATC: <CALLSIGN>, Turn Left heading 290 for ILS Approach, report established.</p> <p>AIRCRAFT: <CALLSIGN>, left heading 290 will report established, maintaining 3000 feet now.</p> <p>AIRCRAFT: <CALLSIGN>, Maintaining 3000 feet established on the ILS now.</p> <p>ATC: <CALLSIGN>, Continue ILS approach, contact Tower now 118.6</p> <p>AIRCRAFT: <CALLSIGN> Continuing ILS approach contacting Tower 118.6</p>	<p>Click right-hand level indicator and choose advised level</p> <p>Click right-hand heading indicator and choose advised heading</p> <p>Click call-sign choose TRANSFER</p>

Standard Instrument Departure Routes

Standard Arrival Routes

Appendix G. Resolution Aircraft and Manoeuvre Selector (RAMS) Procedure (Erzberger, 2006)

1. Step Altitude

Step altitude is a resolution manoeuvre for manoeuvring aircraft in cruise phase (vertical manoeuvre). It is defined by two parameters which are h_{step} and t_r . For each manoeuvre trial, h_{step} is 1000 feet. The definitions of other symbols are given as follows:

t_n = time when the conflict starts

t_{cl} = time when the conflict ends

t_r = time to return to the initial position

$t_r = t_{cl} + 0,5 \text{ minute}$

2. Temporary Altitude

Temporary altitude is a resolution manoeuvre for manoeuvring aircraft in climbing or descending phases (vertical manoeuvre). Similar with step altitude, temporary altitude is also defined by two parameters or h_{step} and t_r . Vertical manoeuvre for each trial is 1000 feet separation (or 1 flight level (FL)).

t_{c1} = time to reach the designated position at the first trial

3. Path Stretch

Path stretch is a lateral manoeuvre which based on time domain resolution algorithm. The basic principle is that any trajectory change for either aircraft that produces a sufficient time shift near the conflict point may potentially prevent the occurrence of conflict.

$$T_{ps} = T_r + \Delta T$$

$$V_{ag} = D_{sr} / T_r$$

$$TR_l = R_l / V_{ag}$$

$$\Delta T_c = \Delta T + T_r - T_{Rl}$$

$$\delta_{ps} = \Delta T_c \cdot V_{ag}$$

$$\delta_{ps} = R_v + R_r - R_l$$

$$R_r^2 = R_v^2 + R_l^2 - 2R_v R_l \cos \psi_v$$

$$R_v = \frac{R_l + \delta_{ps} / 2}{1 + \frac{R_l}{\delta_{ps}} (1 - \cos \psi_v)}$$

$$\psi_v = 15, 30, 45, 60$$

Where:

T_{ps} = the delayed arrival time at the point P_r

T_r = the time for the chosen manoeuvre aircraft to fly from P_s to the return waypoint at P_r

ΔT = the delay for path stretch manoeuvre

D_{sr} = the length of path from P_s to P_r

V_{ag} = the average ground speed along D_{sr}

R_1 = the direct route from P_s to P_r

TR_1 = the time for aircraft A to fly from P_s to P_r via R_1

ΔT_c = the corrected delay time

δ_{ps} = the increment of path length

R_v = the new manoeuvre path

R_r = the new manoeuvre path

Ψ_v = the vector angle for the manoeuvre

Appendix H. the Generated Conflict Encounters

There are two pictures for the each conflict encounter. The first picture shows a current situation. The second picture shows a look-ahead situation when the conflict occurs. The resolution is generated by applying a step altitude, a temporary altitude, and a path stretch methods by Erzberger (2006).

For each secondary conflict encounter due to a wrong advice, two pictures are provided: the current situation and the secondary conflict encounter if the advice is followed. The full plan view display is provided below. For all of the conflict encounters, only the part of encounters will be shown.

1. The first conflict encounter (SWA478 vs UPS76)

The first conflict encounter occurs between SWA478 which is a cruising arrival aircraft and UPS76 which is in cruise. The two aircraft are at same altitude showing a lateral conflict. This conflict refers to the conflict type of cruise versus cruising arrival more than 5 minutes from top of descent in Erzberger (2006). To resolve this conflict, the arrival aircraft should apply a vertical manoeuvre of temporary altitude and the cruise aircraft should execute a step altitude. Given this scenario, the resolution should be that SWA478 descends and maintains at FL 130 and UPS76 climbs and maintains at FL 150 and after 65 seconds, UPS76 should return to its original altitude which is at FL 140.

2. The second conflict encounter (N579BJ vs CHQ3432)

The second conflict encounter occurs between N579BJ and CHQ3432 in which both are arrival aircrafts. The two aircraft are at different altitude (level) but are only separated 500ft vertically. Looking at the full plan view display, this conflict refers to the conflict type of arrival versus arrival with different arrival fixes in Erzberger (2006). To resolve this conflict, the aircraft farthest from arrival fix or in cruise should apply a

vertical manoeuvre of temporary altitude. Given this scenario, the resolution should be that N579BJ descends and maintains at FL 110.

3. The third conflict encounter (AAL1912 vs SKW888)

The third conflict encounter occurs between AAL1912 and SKW888 in which both are cruising arrival aircraft. The two aircraft are at same altitude (co-altitude). Looking at the full plan view display, this conflict refers to the conflict type of cruising arrival versus cruising arrival with same arrival fix in Erzberger (2006). To resolve this conflict, the leading aircraft should increase its airspeed and the trailing aircraft should decrease its airspeed. Given this scenario, the resolution should be that AAL1912 increases the speed to 250 knots and SKW888 reduces the speed to 210 knots. This conflict was included in Study 1.

4. The fourth conflict encounter (SYC876 vs FRL556)

The primary conflict encounter should occur between N579BJ and FRL556 in which both aircraft are cruising. The two aircrafts are at same altitude (co-altitude). This conflict refers to the conflict type of cruise versus cruise with large crossing angle in Erzberger (2006). To resolve this conflict, the aircraft farthest from airspace boundary should apply step altitude and the other aircraft should apply step altitude too. Given this scenario, the resolution should be that FRL556 climbs and maintains at FL 135 and N579BJ descends and maintains at FL 115, and after 60 seconds, those should return to their original altitude.

However, if this advice is followed, there will be a secondary conflict triggered within four minutes after the advice is implemented. The secondary conflict will occur between FRL556 and SYC876. This conflict was included in Study 1.

5. The fifth conflict encounter (SNP567 vs N80866)

The primary conflict encounter should occur between N80866 and SNP567 in which N80866 is a cruising arrival aircraft and SNP567 is a cruising aircraft (has already reached cruise phase). The two aircrafts are at same altitude (co-altitude) when the conflict is projected to occur. This conflict refers to the conflict type of cruise versus cruising arrival more than 5 minutes from top of descent in Erzberger (2006). To resolve this conflict, the arrival aircraft should apply a temporary altitude and the cruise aircraft should apply a step altitude. Given this scenario, the resolution should be that SNP567 climbs and maintains at FL 150 and after the conflict point, the aircraft should return to its original altitude. Meanwhile N80866 should descend and maintain at FL 130, preparing to descend to arrival fix.

However, if this advice is followed, there will be a secondary conflict triggered within two minutes after the advice is implemented. The secondary conflict will occur between N80866 and FRL784 where both aircraft co-altitude. This conflict was included in Study 1.

6. The sixth conflict encounter (CHQ3273 vs CHQ3038)

The sixth conflict encounter occurs between CHQ3273 which is in cruise and CHQ3038 which is a climbing aircraft. The two aircraft are at different altitude showing vertical conflict. This conflict refers to the conflict type of cruise versus climb in Erzberger (2006). To resolve this conflict, the climbing aircraft should apply a vertical manoeuvre of temporary altitude and the cruise aircraft should execute a step altitude. Given this scenario, the resolution should be that CHQ3038 climbs and maintains at FL 119 and CHQ3273 climbs and maintains at FL 140. After 65 seconds, CHQ3038 should continue its climb until reach the exit altitude at FL 140 and CHQ3273 should return to its original altitude which is at FL 130. This conflict was included in Study 1.

7. The seventh conflict encounter (LCL208 vs STI975)

The primary conflict encounter should occur between LCL208 which is a climbing aircraft and STI975 which is a descending arrival aircraft. The two aircraft are at different altitude showing a combination of vertical and lateral conflict. This conflict refers to the conflict type of climb versus descending arrival in Erzberger (2006). To resolve this conflict, the climbing aircraft should apply a vertical manoeuvre of temporary

altitude and the arrival aircraft should execute a path stretch. Given this scenario, the resolution should be that LCL208 climbs and maintains at FL 053 (1000ft before the conflict altitude). After 43 seconds, LCL208 should continue its climbing until reach the exit altitude of FL 140 and maintain at this level. At the same time, STI975 should continue its descent to FL 026 (using an ILS approach) while applying fly-heading to 060 (along the path R_v). After 50 seconds, STI975 should fly heading to 300 (along the path R_r) toward the return waypoint, Pr .

However, if STI975 implements the advice, there will be a secondary conflict triggered within the next 2 minutes between NWA875 and STI975. At the secondary conflict point, STI975 is descending; NWA875 is cruising and the vertical separation between the two aircrafts is violated. The triggered secondary conflict is illustrated in the figure above (on the right).

8. The eighth conflict encounter (NWA875 vs DAL249)

The primary conflict encounter should occur between DAL249 and NWA875 in which both aircraft are cruising. The two aircraft are at same altitude (co-altitude). This conflict refers to the conflict type of cruise versus cruise with large crossing angle in Erzberger (2006). To resolve this conflict, the aircraft farthest from airspace boundary should apply a step altitude and the other aircraft should apply a step altitude as well (both aircraft apply step altitudes). Given this scenario, the resolution should be that DAL249 climbs and maintains at FL 150 and NWA875 descends and maintains at FL 130, and after 65 seconds, those aircrafts should return to their original altitude.

However, if this advice is followed, there will be a secondary conflict triggered within four minutes after the advice is implemented. The secondary conflict will occur between DAL249 and LCL208 since both aircraft co-altitude.

9. The ninth conflict encounter (JIA292 vs CYS753)

The ninth conflict encounter occurs between JIA292 which is a climbing aircraft and CYS753 which is a cruising arrival aircraft. The two aircrafts are at different altitude showing a combination of vertical and lateral conflict. This conflict refers to the conflict type of climb versus cruising arrival in Erzberger (2006). To resolve this conflict, the climbing aircraft should apply a lateral manoeuvre (path stretch) and the arrival aircraft should execute a temporary altitude. Given this scenario, the resolution should be that CYS753 descends and maintains at FL 115. After the conflict point, CYS753 should

continue its descent until reach the arrival fix of FL 026. At the same time, JIA292 should climb and maintain at FL 140 (exit altitude) while applying fly-heading to 260. After 60 seconds, JIA292 should fly heading to 190 to return to its path. This conflict was included in Study 1.

10. The tenth conflict encounter (XAA403 vs FLL234)

The primary conflict encounter will occur between XAA403 and FLL234 in which both aircraft are climbing. The two aircrafts are at different altitude. This conflict refers to the conflict type of climb versus climb in Erzberger (2006). To resolve this conflict, both aircraft should apply a temporary altitude. Given this scenario, the resolution should be that XAA403 climbs and maintains at FL 130 and FLL234 should climb and maintain at FL 120, and after 150 seconds from maneuver issuance, those aircraft should continue climbing to their designated altitude.

Appendix I. Power Analysis Table (Maxwell & Delaney, 2004)

TABLE 13.10
MINIMUM SAMPLE SIZE NEEDED TO ACHIEVE POWER OF .80 WITH $\alpha = .05$

Number of Levels	Minimum Correlation	Design*	<i>d</i>					
			0.25	0.50	0.75	1.00	1.25	1.50
2	—	CRD	253	64	29	17	12	9
	0	RMD	254	65	30	18	13	10
	.1		228	59	28	17	12	9
	.2		203	53	25	15	11	8
	.3		178	46	22	14	10	8
	.4		153	40	19	12	9	7
	.5		128	34	16	10	8	6
	.6		103	28	14	9	7	6
	.7		78	21	11	7	6	5
	.8		53	15	8	6	5	4
3	.9		28	9	6	4	4	3
	—	CRD	310	79	36	21	14	10
	0	RMD	312	81	38	23	16	12
	.1		281	73	34	21	15	11
	.2		250	65	31	19	14	11
	.3		219	58	28	17	12	10
	.4		188	50	24	15	11	9
	.5		158	42	21	13	10	8
	.6		127	34	17	11	9	7
	.7		96	27	14	10	8	6
4	.8		65	19	11	8	6	6
	.9		34	11	7	6	5	5
	—	CRD	350	89	40	23	15	11
	0	RMD	353	92	43	26	19	14
	.1		318	83	39	24	17	13
	.2		284	74	36	22	16	13
	.3		249	66	32	20	15	12
	.4		214	57	28	18	13	11
	.5		179	48	24	16	12	10
	.6		144	39	20	13	10	9
	.7		109	31	16	11	9	8
	.8		74	22	13	9	8	7
	.9		39	13	9	7	6	6