CrossMark

CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Integration of conflict resolution automation and vertical situation display for on-ground air traffic control operations

Fitri Trapsilawati,[1,2] Chun-Hsien Chen,[1]* Chris D. Wickens,[3] and Xingda Qu[4]

[1] School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore.

[2] Department of Mechanical and Industrial Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia.

[3] Colorado State University and Alion Science and Technology, Louisville, CO, USA.

[4] Institute of Human Factors and Ergonomics, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China.

*Corresponding author. E-mail: mchchen@ntu.edu.sg

Abstract
Both conflict resolution aid (CRA) and vertical situation display (VSD) systems may contribute to air traffic control (ATC) operations. However, their effectiveness still needs to be examined before being widely adopted in ATC facilities. This study aims to examine empirically the use of CRA and VSD as well as the systems' interaction in ATC operations. It was found that CRA benefited conflict resolution performance by 13·7% and lowered workload by 46·4% compared with manually performing the task. The VSD could also reduce the air traffic controllers' (ATCOs) workload and improve their situation awareness. Ultimately, when the first CRA failure occurred, the situation awareness supported by VSD offset the performance decrements by 30%. The findings from this study demonstrate that integrating VSD with CRA would benefit ATC operations, regardless of the CRA's imperfection.

1. Introduction

The rapid increase in air traffic demand has become a major challenge for air traffic control (ATC) worldwide, as indicated by a 6·5% increase in traffic demand from 2014 to 2015 (IATA, 2016). This rise presents a challenge for air traffic controllers (ATCOs) in maintaining separation between aircraft. Due to this increase, the probability of air traffic conflicts is also higher, thus imposing higher burdens on ATCOs.

Furthermore, the current ATC systems are approaching their maximum capacity, therefore development of new concepts for the future is even more urgent. The concept of automation of conflict resolution led to the conflict resolution assistant (CORA), in which a tactical plan in the event of conflict is provided to ATCOs to be acknowledged and implemented or rejected (Ehrmanntraut, 2010; SESAR, 2012). However, such automation is still under research and has yet to be employed in active service (SESAR, 2015).

In this study, a conflict resolution aid (CRA) was used to provide advisories to ATCOs specifically on how to manoeuvre aircraft to avoid a potential traffic conflict. However, CRA for ATC operations remains in the development stage because of the high number of possible complex permutations of different flight situations (Kuchar and Yang, 2000; Leone, 2009; SESAR, 2015). Flight situations may be affected by diverse factors such as horizontal situations including overtaking, crossing, converging, and opposite-heading as well as vertical positions covering climb, descend, and level-off. The manoeuvring

dimensions can also vary from speed adjustment to lateral, vertical and combined manoeuvres (Kuchar and Yang, 2000). Because of this, the CRA may not always offer a successful resolution; that is, its reliability is imperfect. Imperfect automation is defined in its reliability level which represents the ratio of successful automation performance to total runs (Rovira et al., 2007; Wickens and Dixon, 2007).

Trapsilawati et al. (2015, 2016a) nevertheless found that such an imperfect (80% reliable) CRA still supported ATCO conflict avoidance performance, well above the level of unaided manual task performance, and similar findings have been observed with other imperfect automation aids (Wickens and Dixon, 2007). In general, conflict resolution automation has gained positive responses (Martin and Imbert, 2012) if the tool does not adversely affect ATCOs' situation awareness (SA) (Kirwan and Flynn, 2002).

Due to the rapid increase in air traffic, the integrated inferences of the three dimensions (i.e. lateral, longitudinal and vertical) have become increasingly critical (Murphy et al., 2012). Furthermore, ATCOs often prefer vertical resolution manoeuvres due to the expediency of the resolution manoeuvre (Erzberger, 2006; Rantanen and Wickens, 2012). Hence, ATCOs must examine the trends of the traffic above and below the potential conflict to assure that those traffic aircraft will not be in the path of the avoidance manoeuvre.

It is also apparent that in nearly all ATC facilities aircraft altitude is only contained in digital data tags, not in the more intuitive spatial displays, and only a few ATC facilities have implemented vertical information-related display. A small number of empirical evaluations of a vertical situation display (VSD) have been conducted. The highly interactive problem solver (HIPS) altitude view (Jorna et al., 1999), level-assessment display (LAD) (SESAR, 2013) and vertical aid window (VAW) (Dehn et al., 2007), for instance, are graphical displays that show the aircraft's predicted climb and descent profiles. VAW was found to maintain SA and assist tactical de-confliction as well as coordination and transfer (Dehn et al., 2007). Another display that also supported the vertical information was WHEELIE, allowing ATCOs to filter the aircraft at a specific flight level by scrolling the Operational Display System-mouse wheel (EUROCONTROL, 2008). It did not, however, depict the vertical plane in graphics format. Our review of the literature fails to reveal any studies in which controller task performance using these VSD concepts has been compared with conventional two-dimensional (map/radar) displays (in an experimental design with high statistical power).

Furthermore, the VSD design in this study improved on prior VSD designs by adding several features. The VSD depicts all waypoints embedded along an aircraft route on the VSD, enhancing the visibility that was not provided in the other tools, to improve the perceptual-cognitive linkages (Woods, 1984) between horizontal and vertical awareness. Next, the VSD in this study enhanced the other tools in terms of the vertical trend information. In the VSD, the prediction of vertical trend information is provided comprehensively with the embedded timing information about when aircraft will be at particular route points near each waypoint. This is to improve the trend visibility of vertical information (Wickens et al., 2013) thus reducing mental computation (Wickens et al., 2000). Lastly, ATCOs are able to de-clutter the display (by removing the VSD) if they wish. ATCOs could activate the vertical profile for respective aircraft by clicking on the aircraft call-sign and re-clicking the call-sign to remove the information. This feature satisfies the 'detail on demand' principle (Shneiderman and Plaisant, 2005) for VSD to help ATCOs conserve their attentional resources.

VSD is not a new auxiliary tool yet it has not been widely implemented in ATC operations. In this study, VSD was integrated with CRA and the use of both tools was empirically examined for the first time. This integration was investigated to examine whether VSD could support the use of CRA, particularly when the CRA errs. In this study, the role of VSD was explicitly examined in mitigating the costs of imperfect CRA automation. The basis of this prediction lies in the logic proposed by Sebok and Wickens (2017). First, the problematic human response to the failures of imperfect automation lies in the loss of SA of the environment controlled and supervised by automation, a finding well supported by the meta-analysis of Onnasch et al. (2014). Second, VSD will reduce the mental computation for the altitude dimension of the airspace; a point supported by prior research in aircraft traffic displays and in ATC (Nunes and Mogford, 2003; Alexander et al., 2005; Dehn et al., 2007). This positive impact

of VSD in improving SA and mental computation through displaying the information should offset the costs on task performance and workload (Hoff and Bashir, 2015) due to imperfection of the automated CRA in this study, particularly given the ATCOs' preference for vertical manoeuvres to avoid conflicts (Kirwan and Flynn, 2002; Rantanen and Wickens, 2012). In a sense then, VSD is designed to create some level of transparency to the automation (Mercado et al., 2016) to buffer the negative effects of imperfect CRA.

This study has three main objectives. The first main objective is to examine the main effect of CRA to establish its task performance benefits over unaided control; and within CRA conditions, to evaluate the costs of imperfection. Although this issue has been examined previously in the laboratory (Trapsilawati et al., 2015, 2016a), those experiments employed primarily student participants with some ATC training. This study employed only ATC professionals. The second main objective is to examine the overall task performance benefits of VSD, independent of whether CRA automation is perfect or not. Such evaluations appear to be infrequent, as described above. The third main objective of this study is to examine how the presence of VSD would interact with CRA unreliability; that is, whether the presence of a VSD could compensate for any incorrect automation recommendation that occurs when CRA works imperfectly. The following hypotheses were tested.

H1: CRA, even if imperfect, would assist task performance (H1a), reduce workload (H1b) and (H1c) increase SA relative to unaided manual task performance.

H2: VSD would improve task performance (H2a), reduce workload (H2b) and increase SA (H2c).

H3: CRA reliability and VSD support would interact, such that the cost of imperfect CRA on task performance (H3a), workload (H3b) and SA (H3c) would be diminished with the presence of VSD.

2. Methods

2.1. Participants

Twenty ATCOs (13 males and 7 females) participated in this study. Their ages ranged from 24 to 62 years (mean = 31·40 years, SD = 10·75 years). The ATCO participants included tower, approach and en-route controllers from the Civil Aviation Authority of Singapore and the Singapore Air Force. Participants' average work experience was 5·18 years with a standard deviation of 5·88 years. The participants were equally assigned to each condition. A power analysis was performed for main and interaction effects, as suggested by Montgomery (2013). For the main effects of automation condition, display and the interaction effect, the parameter for the operational characteristic curve (Φ) was 2·15, 1·89 and 2·12, respectively. With $\alpha = 0.05$, the statistical power for the automation condition, display and the interaction were around 80%, 84% and 92%, respectively, which reflects sufficiently large power. This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at Nanyang Technological University (IRB-2015-08-009). Informed consent was obtained from each participant.

2.2. Apparatus

2.2.1. ATC simulation setup
A medium fidelity ATC simulator, NLR Air Traffic Control Research Simulator (NARSIM) (Ten Have, 1993) representing the Terminal Radar Control (TRACON) facility of Singapore airspace and adjacent en-route sectors, was used to generate various air traffic scenarios. NARSIM employed the standard instrument departure (SID) and standard arrival routes (STAR) of Singapore airspace.

The experiment setup consisted of one ATCO position and two pseudo-pilot positions. In the ATCO position, there were four screens: two 28·05-inch square format 2 K monitors to display a primary radar and flight data respectively, a 12·1-inch touch-based monitor to display the CRA, and a 22-inch touch-based monitor for the VSD.
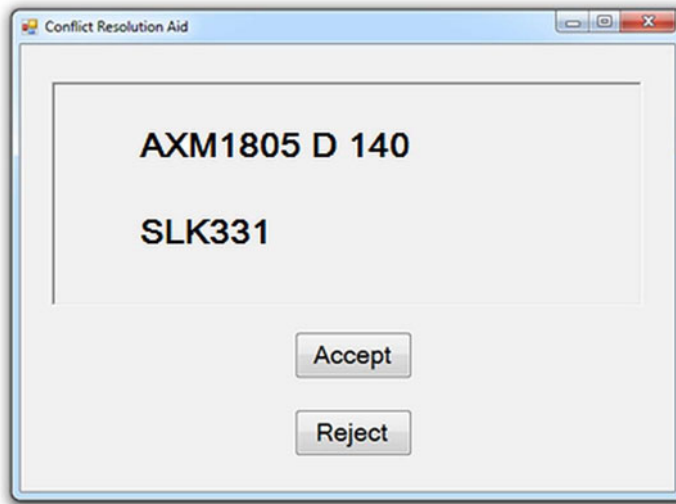
Figure 1. *Example of CRA display.*

In the pseudo-pilot position, there were three screens: two 28·05-inch square format 2 K monitors for the primary radar and the blipper tool to observe the flight status as well as to input manoeuvre commands to an aircraft, and a 12·1-inch touch-based monitor to display the CRA feedback uplinked from the ATCO.

2.2.2. CRA

CRA is an automation aid providing advisories for ATCOs to help resolve impending conflicts. In this study, the advisory was provided in the form of proposed resolution manoeuvres, as shown in Figure 1. The CRA prototype used in this study was developed by Trapsilawati et al. (2015). The CRA worked based on the principles used for a resolution aircraft and manoeuvre selector (RAMS) (Erzberger, 2006). It applied the altitude first resolver principle, where the vertical manoeuvre was suggested first over lateral and speed manoeuvres. The detailed CRA mechanism was described in our previous studies (Trapsilawati et al., 2015, 2016a, 2016b). The list of manoeuvring instructions is provided in Appendix A.

2.2.3. VSD

The VSD used in this study, shown in Figure 2, enhanced the design features of the HIPS, VAW and LAD as described in Section 1. The VSD calculated the aircraft profile based on the air traffic simulation (ATS) scripting in NARSIM. VSD makes use of data from an aircraft task performance library and supports the ATS scripting. The aircraft position was updated by executing a flight from current radar position in NARSIM, following its route, constraints and the script to its destination. This resulted in the vertical and speed profiles which then were delivered to the on-ground system.

Figure 2a shows an example of a predicted conflict between AXM1805 and SLK331. The conflicting aircraft were highlighted in red on the radar display. Both aircraft were at co-altitude as they entered the TRACON airspace with the same arrival fix, with the trailing aircraft travelling faster than the leading aircraft. ATCOs could examine this situation by activating the VSD (Figure 2b). The y and x axes represented the flight level and simulation time, respectively.

ATCOs could also determine whether a secondary conflict would be triggered with the traffic aircraft AWQ8203 (in Figure 2a) that was aiming for approach at the same fix from a different direction. In this case, VSD already provided a clearer position around the merging area that it would pass BOBAG way point at a different altitude and much earlier than the two conflicting aircraft (as shown in Figure 3),
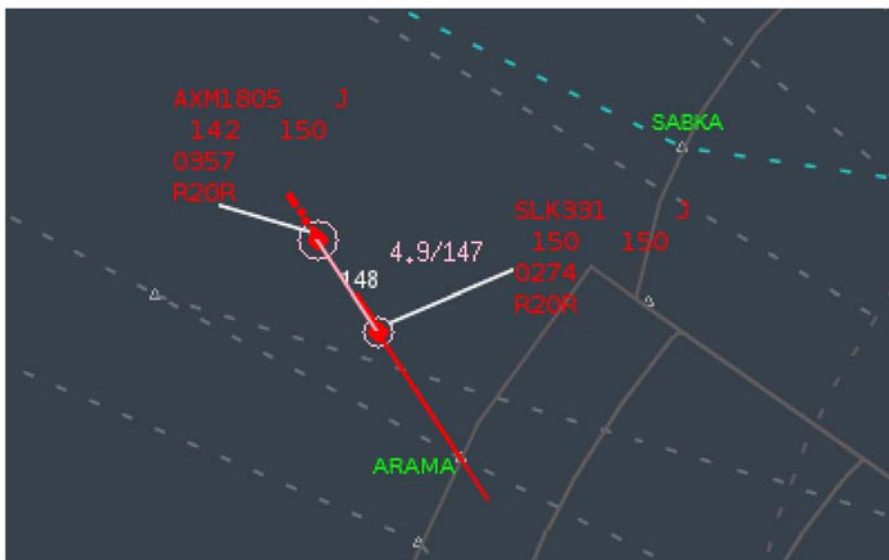
Figure 2. *Example of VSD display: (a) plan view (b) vertical situation display.*

without creating the issue of display ambiguity due to different frame of reference as was the case in HIPS.

In addition, the route points and the vertical trend information were also provided on the VSD, as shown in Figure 2b, where both aircraft were about to descend during the approach phase. Next, from the VSD (Figure 2b), it could earlier be seen that if ATCOs applied vertical separation, the aircraft would not have time to reach the targeted altitude while simultaneously maintaining lateral separation (i.e., 5 nm), as indicated by the close distance of the green triangles showing the way points. This situation could not be accepted since both aircraft were aiming to approach, thus the ATCOs had to instruct 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 manoeuvres.

Figure 3. *Example of conflict in the unreliable automation condition.*

2.3. Experiment design and procedure

The automation condition was varied within-subjects and included three levels: reliable, unreliable and manual conditions. These were counterbalanced between participants. The VSD was a between-subjects factor and was defined by two levels: presence and absence of the VSD, with 10 participants randomly assigned to these two display conditions. The three testing conditions had similar conflict scenarios. However, the traffic patterns were rotated and the aircraft call-signs and waypoints as well as the occurrence times were modified.

In the reliable condition, CRA provided 100% correct manoeuvring advisories for all conflicts. In the unreliable condition, the CRA gave imperfect advice that provided incorrect resolution to the predicted conflict in one conflict out of five (80% CRA reliability). In the manual condition, no CRA was provided; hence, participants performed the ATC tasks manually in this condition.

Participants were provided with a one-hour pre-experiment session that included briefing and training on the ATC simulator, the CRA and VSD. During the experiment session, participants communicated with the pseudo-pilots using voice transmission, and controlled all departing and arriving aircraft within their assigned zones. Participants were required to provide appropriate clearances and maintain separations. While performing the tasks, participants could make use of VSD to supply aircraft vertical information and prediction.

Five pre-set conflicts were placed in each of the testing conditions. Each condition lasted for one hour. The CRA provided a resolution advice two minutes prior to a conflict. Participants could either accept or reject the resolution advice provided by the CRA. If the ATCO accepted the advice, additional information containing the resolution advice would be automatically uplinked to the pseudo-pilot's screen. The pseudo-pilot would directly apply the resolution by executing pre-set commands in the simulator. The commands were set to resolve the conflict and return the aircraft to its flight path as determined by the initial flight plan. If the ATCO rejected the resolution advice, the CRA would stop processing the respective aircrafts' data, thus ATCOs verbally provided their own resolution manoeuvre which would be executed by the pseudo-pilots.

2.4. Dependent measures

2.4.1. Task performance

Conflict resolution task performance was represented by the percentage of resolved conflicts, which was operationally defined as the absence of loss of separation following the CRA's instruction. The

percentage of resolved conflict denotes the ratio of the conflicts resolved by the ATCOs and the total number of air traffic conflicts in the scenario.

2.4.2. *Operator workload*

The subjective workload measurement was obtained using the NASA-TLX (Hart and Staveland, 1988) which was administered upon the completion of each experimental condition. The objective measures of mental workload were assessed using ready response latency and percentage of timeouts in the situation present assessment method (SPAM) (Durso et al., 2004). SPAM ready response latency indicates the readiness of participants to answer SA questions while not knowing what the SA queries were, and often correlates with objective workload (Strybel et al., 2008; Vu et al., 2012). Percentage of timeouts was defined by the ratio of non-responded questions to the total probes. There were nine ready prompts provided throughout the one-hour simulation in each condition.

2.4.3. *SA*

Participants were required to respond to SA probes pertaining to the conflicting aircraft as well as other aircraft in the airspace; the probes appeared every 6 min throughout the one-hour experiment in each testing condition. SA measures were derived from the SA question probes (Durso et al., 2006). The accuracy and time taken to answer the SPAM queries (i.e., SA question probes) reflect SA (Durso et al., 2004). The SA measures included SA probe response latency (i.e., time taken to answer SA questions) and accuracy (i.e., percentage of correct responses to the total SA probes).

2.5. *Analysis*

The data were first examined for outliers and adjusted using using Inter Quartile Range rule in case of normality violation. Afterwards, 3(automation condition) $\times$ 2(display) mixed-design repeated-measure analysis of variance (ANOVA) was conducted for all measures. The alpha level was set at 0·05. The assumption of sphericity was also tested using Mauchly's sphericity test (Montgomery, 2013). Greenhouse-Geisser adjustment was adopted for sphericity violation. Moreover, post-hoc tests using Least Significant Difference (LSD) were conducted for significant main effects. To further analyse the correct versus failed automation trials within the unreliable block, proportion testing was performed to compare the first automation failure trial with the correct automation trials.

3. Results

3.1. *Task performance: percentage of resolved conflicts*

To test Hypothesis 1a, the mixed repeated-measure ANOVA showed a significant effect of the automation condition, $F(1·31, 28) = 7·70$, $P = <0·01$, $\eta_p^2 = 0·819$). Figure 4 shows the significant monotonic increase in performance from manual to unreliable and reliable automation. Post-hoc test using LSD test showed that the task performance in the reliable condition ($M = 100·0$, $SD = 0·00$) was significantly better than in the manual condition ($M = 86·3$, $SD = 15·86$) ($P = <0·01$). This indicated that the minimum separation was infringed in around 14% of cases when ATCOs manually resolved conflicts. There was no significant difference between reliable and unreliable ($M = 96·3$, $SD = 8·06$) ($P = 0·102$) nor between unreliable and manual ($P = 0·031$). Hypothesis 1a was therefore partially confirmed.

There was a trend that VSD did improve the task performance ($M_{VSD} = 96·2$, $SD_{VSD} = 3·56$ vs $M_{NoVSD} = 92·59$, $SD_{NoVSD} = 9·67$), however this effect was not significant $F(1, 14) = 1·27$, $P = 0·278$, $\eta_p^2 = 0·083$). Therefore, Hypothesis 2a was not confirmed.

The general interaction of VSD and automation condition was not significant, $F(1·31, 28) = 0·44$, $P = 0·569$, $\eta_p^2 = 0·102$. However, to further see the effect of VSD during off-nominal situation, the test of proportions was performed between all the correct automation trials versus the failed automation trial in the unreliable block for each display condition, as shown in Figure 5. The results revealed that the task performance in correct and failed automation trials did not differ significantly (0%) when VSD
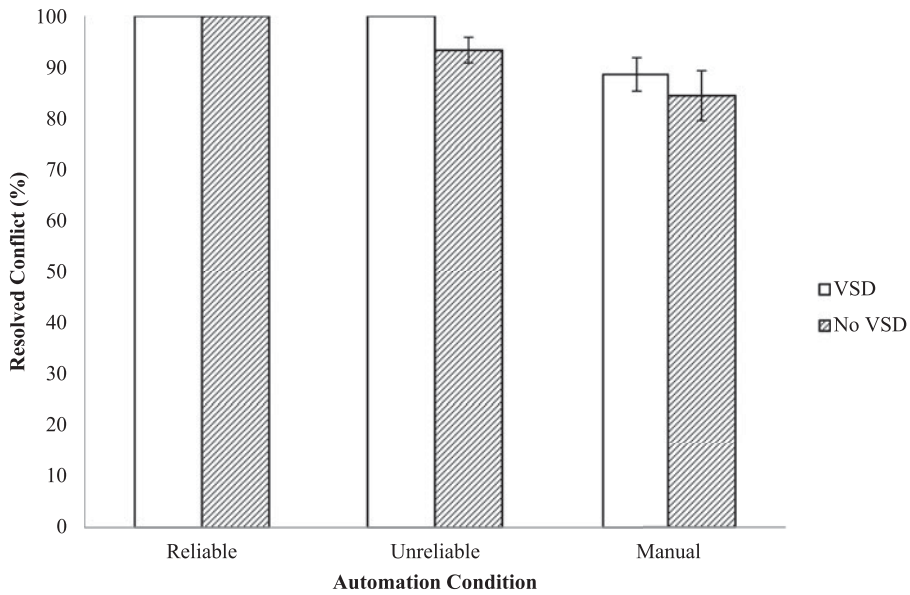
Figure 4. *Conflict resolution performance (error bars indicate 1 standard error (SE)).*
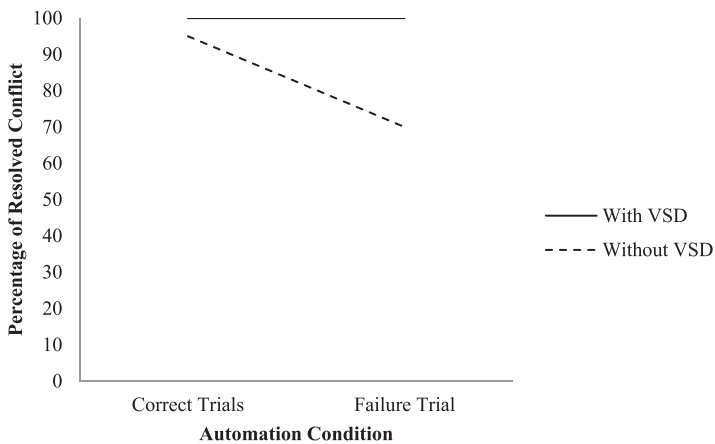

Figure 5. *First-failure effect.*

was present, $Z = 0.00$, $P = 1.00$, $r = 1.00$. However, the task performance in the automation failure trial was significantly worse than the automation correct trials, by 25% when VSD was absent, $Z = 2.36$, $P = 0.018$, $r = 0.33$. Therefore, Hypothesis 3a was confirmed; VSD could diminish the cost of automation imperfection as shown by 0% loss in performance when the automation erred.

3.2. Mental workload

To test Hypothesis 1b, 3×2 mixed repeated-measure ANOVA analyses were performed for both subjective workload and objective workload measures (percentage of timeouts and ready response latency). The results showed that the main effect of CRA on subjective workload was not significant, $F(2, 30) = 0.28$, $P = 0.76$, $\eta_p^2 = 0.018$. However, the main effect of CRA on objective workload was significant, as indicated by percentage of timeouts as shown in Figure 6, $F(2, 30) = 4.39$; $P = 0.021$; $\eta_p^2 = 0.23$. The post-hoc tests (LSD) showed that workload with the reliable CRA ($M = 10.4$, $SD = 12.25$)
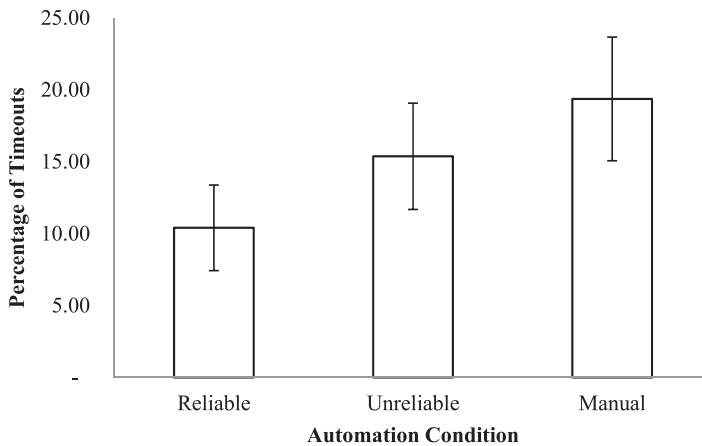
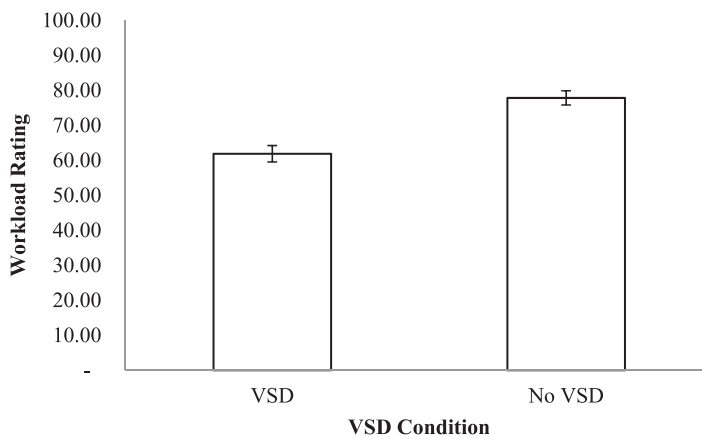**Figure 6.** *Percentage of timeouts (error bars indicate 1 SE).*

Figure 7. *Workload rating (error bars indicate 1 SE).*

was lower than manual ($M = 19 \cdot 4$, $SD = 17 \cdot 73$) ($P = 0 \cdot 016$). No significant differences were found between reliable and unreliable ($M = 15 \cdot 4$, $SD = 15 \cdot 22$) ($P = 0 \cdot 073$) nor between unreliable and manual ($P = 0 \cdot 26$). Hypothesis 1b was therefore partially confirmed since workload with the imperfect CRA was not different in the manual condition. For ready response latency, the main effect of the automation condition was not significant, $F(1 \cdot 46, 20 \cdot 4) = 0 \cdot 42$, $P = 0 \cdot 60$, $\eta_p^2 = 0 \cdot 029$.

Regarding Hypothesis 2b, subjective workload (Figure 7) was significantly lower with VSD ($M = 62 \cdot 6$, $SD = 9 \cdot 73$) than without it ($M = 77 \cdot 8$, $SD = 8 \cdot 37$), $F(1, 15) = 26 \cdot 9$, $P = {<}0 \cdot 01$, $\eta_p^2 = 0 \cdot 64$. As with subjective workload, the percentage of timeouts was lower with VSD; however the effect did not reach conventional levels of statistical significance, $F(1, 15) = 3 \cdot 28$, $P = 0 \cdot 09$, $\eta_p^2 = 0 \cdot 18$. In terms of response latency, the effect of VSD on objective workload was also significant, as indicated by lower ready response latency with VSD ($M = 5 \cdot 38$ s, $SD = 2 \cdot 50$ s) than without it ($M = 7 \cdot 78$ s, $SD = 2 \cdot 72$ s), $F(1, 14) = 9 \cdot 46$, $P = 0 \cdot 008$, $\eta_p^2 = 0 \cdot 40$. These findings on subjective workload and ready response latency therefore indicated that Hypothesis 2b was confirmed.

For Hypothesis 3b, no significant interaction effect was found on subjective workload, $F(2, 30) = 0 \cdot 62$, $P = 0 \cdot 54$, $\eta_p^2 = 0 \cdot 040$, on percentage of timeouts, $F(2, 30) = 1 \cdot 65$, $P = 0 \cdot 21$, $\eta_p^2 = 0 \cdot 099$, nor on ready response latency, $F(1 \cdot 46, 20 \cdot 4) = 0 \cdot 51$, $P = 0 \cdot 55$, $\eta_p^2 = 0 \cdot 04$. These findings showed that Hypothesis 3b was not supported. Collectively, the findings are conclusive that VSD reduced workload in all conditions.
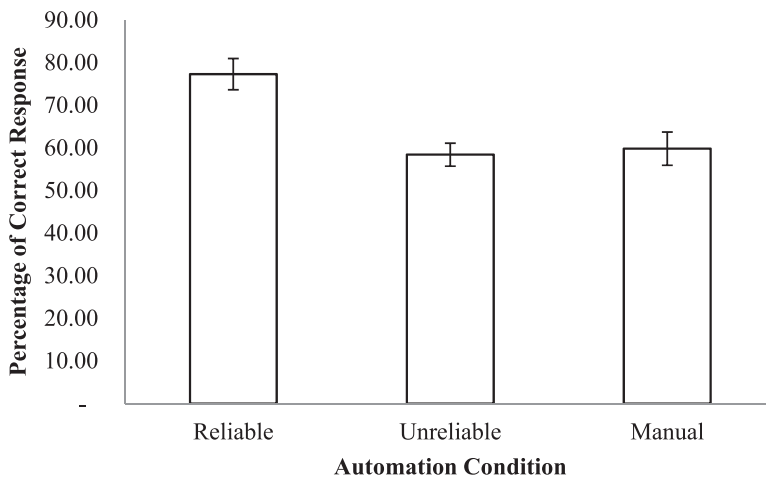
Figure 8. *Percentage of correct response (error bars indicate 1 SE).*

3.3. SA

The results shown in Figure 8 revealed a significant influence of automation condition on SA as indicated by SA probe accuracy $F(2, 36) = 12·24$, $P = <0·01$, $\eta_p^2 = 0·40$. SA was substantially improved by reliable CRA ($M = 77·3\%$, $SD = 21·3\%$) compared with unreliable CRA ($M = 58·4\%$, $SD = 12·6\%$) ($P = 0·001$) and manual condition ($M = 59·8\%$, $SD = 20·4\%$) ($P = <0·01$), but SA probe accuracy did not differ between the latter two conditions ($P = 0·78$). Therefore, these findings partially supported Hypothesis 1c. Regarding SA probe response latency, no significant effect of automation condition, $F(2, 30) = 2·97$, $P = 0·066$, $\eta_p^2 = 0·17$ was found.

For Hypothesis 2c, the findings showed the SA probe accuracy was significantly higher with VSD ($M = 72\%$, $SD = 10·0\%$) than without it ($M = 58·3\%$, $SD = 20·6\%$), $F(1, 18) = 5·33$, $P = 0·033$, $\eta_p^2 = 0·23$. This finding thus showed that Hypothesis 2c was supported. The effect of VSD on SA probe response latency was not significant, $F(1, 15) = 0·10$, $P = 0·76$, $\eta_p^2 = 0·006$.

Regarding Hypothesis 3c, the results revealed that there was no interaction effect found on the SA probe accuracy, $F(1·34, 21·5) = 0·26$, $P = 0·77$, $\eta_p^2 = 0·016$. However, there was a significant interaction effect between automation condition and VSD on SA probe response latency (Figure 9), $F(2, 30) = 4·09$, $P = 0·027$, $\eta_p^2 = 0·21$. This interaction reveals that in the reliable condition, the probe response latency without VSD was shorter than with it, $t(18) = 2·71$, $P = 0·014$, but this difference was not even close to significance in the unreliable condition, $t(17) = 0·20$, $P = 0·84$, and manual condition, $t(17) = 0·47$, $P = 0·65$, partially confirming Hypothesis 3c.

4. Discussion

The discussion below examines, in turn, the main effects of automation level (H1), the benefits of VSD (H2), and the interaction between these two (H3), each as represented in our three hypotheses.

4.1. Automation effect

While overall performance in resolving conflicts did not significantly degrade when the automation became imperfect, partially supporting Hypothesis 1a, Figure 5 shows vividly that performance *did* suffer in the infrequent trials when automation failed (when VSD was not available). We can also observe no significant cost from imperfect automation compared with the manual performance. This is in contrast to the finding of Metzger and Parasuraman (2005), who found that imperfect automation
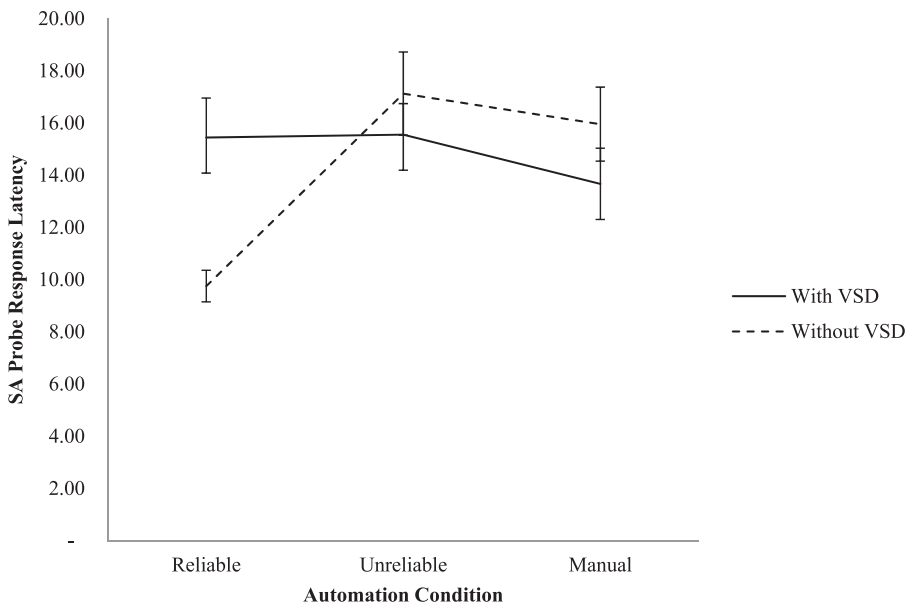
Figure 9. *Probe response latency (error bars indicate 1 SE).*

led to worse task performance than in manual condition. This different result might be explained by the different nature of the tasks (i.e., detection versus resolution).

Finally, beyond the overall performance effects, it is clear that perfect CRA automation supported the two secondary variables of critical importance in system design: reducing workload (see Figure 6) (Young et al., 2015) as compared with manually performing the task, partially supporting Hypothesis 1b, and improving SA accuracy (see Figure 8) (Stanton et al., 2017), showing that Hypothesis 1c was upheld. The latter was significant in the performance-based measure of workload, but not in subjective workload.

4.2. Benefits of VSD

Implementation of VSD was clearly of benefit. Although it did not significantly improve performance over all trials (showing that Hypothesis 2a was not upheld – but it trended in that direction, see Figure 5), its benefit was clearly seen both in the reduction of subjective workload (Figure 7) and in the increase in the accuracy of SA (Figure 8) (Corver et al., 2016), confirming Hypotheses 2b and 2c. The ATCOs' workload was reduced and their SA increased with VSD, as indicated by the higher response accuracy which we also found in previous studies (Trapsilawati et al., 2017; Trapsilawati and Chen, 2017). These findings were in line with those reported in Rovira et al. (2007), that operators' performance is generally better when they are provided with the contextual information supporting their mental model. Comparing the reliable conditions in Figure 8 (accuracy) and Figure 9 (speed), we would argue that the 25% gain in accuracy achieved with VSD more than offsets the 6 s slowing in response latency; and of course such a speed–accuracy trade-off was not present at all in the other two conditions, partially supporting Hypothesis 3c.

4.3. VSD to offset the costs of automation imperfection

VSD designed specifically to assist in the compensation of improved SA can restore the performance loss of automation imperfection. While the loss in overall performance with imperfect automation was sufficiently small that this compensation was not revealed statistically (Figure 4), when our attention

was specifically focused on the automation failure trials alone, the hypothesised interaction was strong (Figure 5), supporting Hypothesis 3a. In the failure trial, the transparency of the automation offered by VSD significantly improved the ATCOs' performance and eliminated the decrement that the failure had caused. Hypothesis 3b was not upheld since no interaction effect was found on the workload measures. This finding might be explained from the fact that VSD could lower ATCOs' subjective and objective workload (i.e., lower ready response latency) regardless of the CRA conditions.

Our findings have important implications for air traffic management, as well as for the science of human–automation interaction. First, the benefits of CRA continue to suggest that this is a promising form of automation support tool for the ATCO (Prevot et al., 2012; Trapsilawati et al., 2015, 2016a). This benefit was confirmed not only from the empirical data obtained in this study but also from the feedback from the ATCO participants which was favourable to the CRA. The ATCOs indicated that the CRA was helpful to support their conflict resolution work. However, we would argue that this benefit would only be realised: (1) as long as the raw data of the traffic picture are readily available to the ATCO (and here, the benefits were enhanced when more raw data were provided by the VSD) and (2) the automation of decision aiding is at a lower level, such that automation only recommends a manoeuvre (as was the case in this study), but does not automatically implement it (Parasuraman et al., 2000; Onnasch et al., 2014).

Second, our findings about the benefits of VSD suggest a promising technology for assisting ATCOs in their task performance. We found that the reductions in workload and improvement in SA provided by the display were pronounced. This finding was also supported by the ATCO participants' feedback, revealing that they could examine the vertical situation using the VSD although they needed sometimes to get used to comprehending it. 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 is parallel to that found by continuous predictive display visualisations in the process control industry (Yin et al., 2015). Furthermore, the ATCOs' performance with VSD was descriptively better than without it (Figure 4), even if it did not improve direct task performance overall. Despite the trend of better performance with VSD, perhaps the difference failed to reach significant level due to the overall high performance of the professional ATCO participants.

There were some limitations to this study. First, the partial fidelity of the ATC simulator used in this study may not entirely reflect the real situation since some factors, such as weather, were not implemented in the study. ATC simulation has been utilised in previous research and generated appropriate results because it is almost impossible to test a new safety-critical future-functionality concept in the real-world system (Prevot et al., 2012). However, the weather factor should be considered in future research. Second, ATCO participants varied from tower to en-route ATCOs. In future research, the participants should only consist of TRACON ATCOs who are more familiar with the TRACON area. Third, the experiment could be further extended to participation of more ATCOs and different conditions of faulty automation, as well as longer experiment time to better reflect ATC operations.

References

Alexander, A. L., Wickens, C. D. and Merwin, D. 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.

Corver, S. C., Unger, D. and Grote, G. (2016). Predicting air traffic controller workload trajectory uncertainty as the moderator of the indirect effect of traffic density on controller workload through traffic conflict. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **58**(4), 560–573.

Dehn, D., Lowe, C. and Hill, C. (2007). *First ATC Support Tools Implementation (FASTI): Cognitive Task Analysis*. Brussels: EUROCONTROL. Available at: https://www.eurocontrol.int/sites/default/files/article/content/documents/nm/fasti-hf-cognitive-task-analysis2007.pdf

Durso, F. T., Dattel, A. R., Banbury, S. and Tremblay, S. (2004). SPAM: The real-time assessment of SA. In Banbury, S., Tremblay, S. (eds.). *A Cognitive Approach to Situation Awareness: Theory and Application*. (Vol. **1**, 137–154). Hampshire, UK: Ashgate.

Durso, F. T., Bleckley, M. L. and Dattel, A. 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.

Ehrmanntraut, R. (2010). Full automation of air traffic management in high complexity airspace. Ph.D. thesis, Technical University of Dresden, Germany.

Erzberger, H. (2006). Automated Conflict Resolution for Air Traffic Control. *Paper Presented at the 25th International Congress of the Aeronautical Sciences for the Society of International Council of the Aeronautical Sciences,* 3–8 September. Hamburg, Germany.

EUROCONTROL. (2008). *Wheelie – Advanced Display Filtering Technique (Human Factors Experiment)*. Brussel: EUROCONTROL. Available at: http://www.eurocontrol.int/sites/default/files/library/005_WHEELIE_advanced_display_filtering.pdf.

Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. *Human Mental Workload*, **1**(3), 139–183.

Hoff, K. A. and Bashir, M. (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.

IATA. (2016). Demand for air travel in 2015 surges to strongest result in five years. Available at: http://www.iata.org/pressroom/pr/Pages/2016-02-04-01.aspx.

Jorna, P. G. A. M., Pavet, D., van Blanken, M. and Pichancourt, I. (1999). PHARE Ground Human Machine Interface (GHMI) Project: Summary Report. Brussels: EUROCONTROL. Available at: https://www.eurocontrol.int/phare/gallery/content/public/documents/99-70-02ghmi.pdf

Kirwan, B. and Flynn, M. (2002). CORA 2 Investigating Air Traffic Controller Conflict Resolution Strategies. Brussels: EUROCONTROL. Available at: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=B0A071A8FA6B6065572B8DA09572824B?doi=10.1.1.80.2523&rep=rep1&type=pdf.

Kuchar, J. K. and Yang, L. C. (2000). A review of conflict detection and resolution modeling methods. *IEEE Transactions on Intelligent Transportation Systems*, **1**(4), 179–189.

Leone, M. (2009). Tactical Controller Tool Real Time Simulation Final Report*, Vol. **3**. UK: EUROCONTROL. Available at: http://www.eurocontrol.int/sites/default/files/article/content/documents/nm/fasti-tct-rts-2009.pdf.

Martin, C. and Imbert, J–P. (2012). Introduction of a More Automated Environment in En-Route Air Traffic Control. *Paper Presented at the Second SESAR Innovation Days,* 27–29 November. Braunschweig, Germany.

Mercado, J., Rupp, M., Chen, J., Barnes, M., Barber, D. and Procci, K. (2016). Intelligent agent transparency in human-agent teaming for multi-UxV management. *Human Factors*, **58**, 401–415.

Metzger, U. and Parasuraman, R. (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.

Montgomery, D. C. (2013). *Design and Analysis of Experiments*. 8th Edition. New Jersey: John Wiley & Sons, Inc.

Murphy, E. D., Albert, H. A., Chen, J. M. and Anderson, G. G. (2012). The Role of Mental Computations in Current and Future En Route Air Traffic Control. *Paper Presented at the Proceedings of the 56th Human Factors and Ergonomics Society Annual Meeting*, Boston, MA, USA.

Nunes, A. and Mogford, R. H. (2003). Identifying Controller Strategies that Support the 'Picture'. *Paper Presented at the 47th Annual Meeting for the Human Factors and Ergonomics Society*, 13–17 October. Denver, CO.

Onnasch, L., Wickens, C. D., Li, H. and Manzey, D. (2014). Human task 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, R., Sheridan, T. B. and Wickens, C. 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.

Prevot, T., Homola, J. R., Martin, L. H., Mercer, J. S. and 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.

Rantanen, E. M. and Wickens, C. D. (2012). Conflict resolution maneuvers in air traffic control: investigation of operational data. *The International Journal of Aviation Psychology*, **22**, 266–281.

Rovira, E., McGarry, K. and Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*, **49**(1), 76–87.

Sebok, A. and Wickens, C. D. (2017). Implementing lumberjacks and black swans into model-based tools to support human–automation interaction. *Human Factors*, **59**(2), 189–203.

SESAR. (2012). SESAR Concept of Operations. SESARJU Report. Available at: https://www.eurocontrol.int/sites/default/files/field_tabs/content/documents/sesar/20070717-sesar-conops.pdf.

SESAR. (2013). Final Project Report on the concept and benefits for improving TP using AOC data: Improved Airline Flight Plan Information into ATC Trajectory Prediction (TP) Tool. SESARJU Report. Available at: http://www.sesarju.eu/sites/default/files/solutions/3_AOC_Data_for_TP_Final_Project_Report.pdf?issuusl=ignore.

SESAR. (2015). Automated support for conflict detection, resolution support information and conformance. ESSIP Plan Edition 2015. Available at: https://www.eurocontrol.int/sites/default/files/content/documents/officialdocuments/reports/atc12-1.pdf

Shneiderman, S. B. and Plaisant, C. (2005). *Designing the User Interface*. 4th Edition. Boston, MA, USA: Pearson Addison Wesley.

Stanton, N. A., Salmon, P. M., Walker, G. H., Salas, E. and Hancock, P. A. (2017). State-of-science: situation awareness in individuals, teams and systems. *Ergonomics*, **60**(4), 449–466.

Strybel, T. Z., Vu, K.-P. L., Kraft, J. and Minakata, K. (2008). Assessing the Situation Awareness of Pilots Engaged in Self Spacing. *Paper Presented in the Proceedings of the 52th Human Factors and Ergonomics Society Annual Meeting*, 22–26 September. New York.

Ten Have, J. M. (1993). The development of the NLR ATC Research Simulator (NARSIM): Design philosophy and potential for ATM research. *Simulation Practice and Theory*, **1**(1), 31–39.

Trapsilawati, F., Qu, X., Wickens, C. D. and Chen, C. H. (2015). Human factors assessment of conflict resolution aid reliability and time pressure in future air traffic control. *Ergonomics*, **58**(6), 897–908.

Trapsilawati, F., Wickens, C. D., Qu, X. and Chen, C. H. (2016a). Benefits of imperfect conflict resolution advisory aids for future air traffic control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **58**(7), 1007–1019.

Trapsilawati, F., Chen, C. H. and Khoo, L. P. (2016b). An Investigation into Conflict Resolution and Trajectory Prediction Aids for Future Air Traffic Control. *Proceedings of the 23th ISPE Inc. International Conference on Transdisciplinary Engineering*, 4–6 October. Brazil: Advances in Transdisciplinary Engineering, 503–512.

Trapsilawati, F. and Chen, C. H. (2017). Effects of Information Availability on Workload and Situation Awareness in Air Traffic Control. *Proceedings of the 24th ISPE Inc. International Conference on Transdisciplinary Engineering*, 10–14 July. Singapore: Advances in Transdisciplinary Engineering, 21–28.

Trapsilawati, F., Wickens, C. D., Chen, C. H. and Qu, X. (2017). Transparency and Conflict Resolution Automation Reliability in Air Traffic Control. *Proceedings of the 19th International Symposium of Aviation Psychology*, 8–11 May. Dayton, OH, USA.

Vu, K.-P. L., Strybel, T. Z., Battiste, V., Vernol, L. J., Dao, A.-Q. V., Brandt, S., Ligda, S., *et al.* (2012). Pilot task 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.

Wickens, C. D. and Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, **8**, 201–212.

Wickens, C. D., Gempler, K. and Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors*, **2**(2), 99–126.

Wickens, C. D., Hollands, J. G., Banbury, S. and Parasuraman, R. (2013). *Engineering Psychology and Human Performance 4th Edition*. New Jersey, USA: Pearson.

Woods, D. 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.

Yin, S., Wickens, C. D., Helander, M. and Laberge, J. C. (2015). Predictive displays for process-control schematic interface. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **57**(1), 110–124.

Young, M., Brookhuis, K., Wickens, C. D. and Hancock, P. (2015). State of the science in mental workload. *Ergonomics*, **58**, 1–17.

Appendix A: List of abbreviations for manoeuvres

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