2008

Analysis of human and automated separation assurance at varying traffic levels

Jeffrey R. Homola
San Jose State University

Recommended Citation
Homola, Jeffrey R., "Analysis of human and automated separation assurance at varying traffic levels" (2008). *Master's Theses*. 3575.
DOI: https://doi.org/10.31979/etd.ykd3-u9uf
https://scholarworks.sjsu.edu/etd_theses/3575

ANALYSIS OF HUMAN AND AUTOMATED SEPARATION ASSURANCE AT VARYING TRAFFIC LEVELS

A Thesis

Presented to

The Faculty of the Graduate Program in Human Factors and Ergonomics

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Jeffrey R. Homola

August 2008

ProQuest LLC
789 E. Eisenhower Parkway
PO Box 1346
Ann Arbor, MI 48106-1346

APPROVED FOR THE DEPARTMENT OF INDUSTRIAL

SYSTEMS ENGINEERING

Kevin Jordan, Ph.D.

Thomas Prevôt, Ph.D.

Mark Van Selst, Ph.D.

APPROVED FOR THE UNIVERSITY

Rhea I. Williamson 07/10/08

ABSTRACT

ANALYSIS OF HUMAN AND AUTOMATED SEPARATION ASSURANCE AT VARYING TRAFFIC LEVELS

by Jeffrey R. Homola

As part of the Advanced Airspace Concept, an algorithm for automatically resolving aircraft conflicts has been developed to partially offset air traffic controller workload and provide consistency. This study evaluated the differences in performance between humans resolving conflicts in a manual and interactive mode as well as a fully automated conflict resolution mode. This was done at current day, twice, and three times that level of traffic. Workload impact and acceptability of the algorithm's resolutions were also investigated. Results suggest that the automation provided significant benefits in terms of safety and efficiency particularly at higher levels of traffic. There was also a significant reduction in workload. The resolutions provided by the automation were also rated as being generally acceptable.

ACKNOWLEDGEMENTS

I would first like to thank Dr. Kevin Jordan for his continual patience and guidance throughout the entire thesis process as my advisor. Special thanks also go to Dr. Thomas Prevôt for his inspiration and direction, without which this study would not have been possible. I would also like to thank Dr. Mark Van Selst for his willingness to volunteer his time to serve on this committee following the loss of Dr. Kevin Corker. A heartfelt thank you also goes out to the late Dr. Corker, whose memory will live on in all of the people he has touched.

I would also like to thank Todd Callantine, Paul Lee, Joey Mercer, and Quang Dao for all of their help along the way. I must also thank Ev Palmer and Nancy Smith for providing me with the opportunity to work in such a great environment with such great people.

A special acknowledgement must also go out to Dr. Heinz Erzberger. Without his vision of an advanced airspace concept, this study would not have been possible.

I am eternally indebted to my family for their unconditional love and support throughout my life. Nothing I have ever accomplished could have been possible without the solid foundation that you have provided me.

Finally, to my wife and young son, thank you for giving my life a true purpose. You are a true joy and an inspiration for everything that I do.

TABLE OF CONTENTS

SECTION PAGE

LIST OF FIGURES

INTRODUCTION

By any method of estimation, the National Airspace System (NAS) of the United States is very busy. For example, there are approximately 7,000 aircraft in the air at any given time throughout the day across the United States (Federal Aviation Administration Air Traffic Organization, 2006). This figure translates to roughly 50,000 aircraft transiting the NAS each day. As the airspace is busy, so too are the more than 39,000 men and women that make up the air traffic controller workforce. These individuals are responsible for the safe handling of the roughly 46 million aircraft that the 800 million passengers and over one trillion dollars in cargo depend upon for their travel and business needs each year.

Safety

Despite the sheer volume of air traffic and the inherent danger with which that brings, the air safety record for the United States has remained strong. For instance, the FAA reported that 2005 saw the "lowest airline fatal accident rate in the history of aviation" (Federal Aviation Administration Air Traffic Organization, 2006, p. 10). According to some estimates, an individual could board a domestic flight every day for 36,000 years before being involved in a fatal accident (Barnett, 2001). This consistent level of safety at which the air transportation system operates is in no small part due to the manner in which the air traffic controllers across the country perform their duties.

A System Under Stress

However, there are some ominous signs that the current air traffic system is under stress. Although the fatal accident rates are at an all-time low, other measures of safety paint a different picture. For example, in 2005 there were a total of 681 Category A and B operational errors (Federal Aviation Administration Air Traffic Organization, 2006). An operational error is defined as an error committed by an air traffic controller that results in two aircraft exceeding the defined separation minima of five nautical miles (nm) laterally and 1000 feet (ft) vertically. Related to the cited figure, a Category A error is one in which the error is labeled as being high in severity whereas a Category B error refers to an error that is moderate-uncontrollable in which the controller lacked sufficient time to avoid the error despite being aware of its impending occurrence (Federal Aviation Administration, 2004). Although this number was down from figures reported in 2003, it exceeded the acceptable limit of 637 set by the FAA (Federal Aviation Administration Air Traffic Organization, 2006). Additionally, when taken in the context of a longer time span, these numbers indicate a 68% increase in operational errors when compared to those reported in 1998 (Carroll & Guadiano, 2006).

There are also more benign indications of a system under stress. One indication can be seen through the use of Ground Delay Programs (GDP). Implementation of GDPs is one way of dealing with the problem of over-congestion in that they can be called upon when the volume of traffic directed toward an airport exceeds its handling capacity. In a study of the rate of GDP implementation, researchers found that its use has increased each year since 1998 (Krozel, Hoffman, Penny & Butler, 2003). Delays do not come without associated costs, however. It has been estimated that the average financial cost incurred by delays –not just those resulting from ground delay programs but all conditions that may lead up to its initiation- is 5.9 billion dollars in direct operating costs per year (Air Transport Association, 2006). When viewed in a broader context with respect to commercial and private interests, the impact of delay is estimated to cost upwards of 30 billion dollars per year as a result of people and products not making it to their destinations on time (Joint Planning and Development Office, 2004).

Operational errors and delay are but two examples of problems stemming from the increasingly dense and complex traffic situation in the NAS today. There is no indication that these trends will change any time in the near future either. Using the number of enplanements as an example, domestic figures show that between the years of 2004 and 2005 alone there was a 6.6 increase (Federal Aviation Administration, 2006). Additionally, when compared to earlier numbers, there has been a progressive twofold increase in enplanements since the mid-1970s (Bonnefoy & Hansman, 2005). In fact, according to forecasts made by the FAA, commercial aviation appears to be set to handle 1 billion passengers annually by the year 2015 (Federal Aviation Administration, 2006). This translates to a further estimate that there will be a threefold increase in air traffic by the year 2025. According to a Joint Planning and Development (JPDO) document, the current system is ill-equipped to handle the over-saturation brought on by even a twofold increase (2004).

Looming Air Traffic Control Staffing Crisis

As alluded to earlier, air traffic controllers can be considered the glue that holds the current air transportation system intact. However, even this aspect of the system faces a potential crisis with its origins rooted in the bitter Professional Air Traffic Controllers Organization (PATCO) strike of 1981. In response to the demands set forth by the striking union, former president Reagan summarily fired over 11,000 air traffic controllers. A replacement force was hired in roughly equal numbers within the following decade, which provided a relatively stable workforce for the years that followed. However, with a mandatory retirement age of 56, the mass hirings of the eighties would serve to undermine that stability as the majority of the workforce would become eligible for retirement between the years of 2002 and 2012 (Nolan, 2004). More specifically, a 2004 FAA report stated that 73 percent of the controller workforce would be eligible to retire within the next 10 years. This leads directly to the issue of understaffing at a time when air traffic density levels are concurrently increasing. According to a National Air Traffic Controllers Association (NATCA) article, “even in the best case scenario, the system will be left woefully understaffed for years to come” (National Air Traffic Controllers Association, n.d.).

Recognition and Response to Problem

In response to the looming situation outlined above, there has been a call from all corners and levels of the aviation industry and government for a necessary and decisive change to the existing air transportation system. In a 2004 speech to the Aero Club of Washington, then secretary of transportation Norman Mineta stated that, “The changes that are coming are too big, too fundamental for incremental adaptations of infrastructure.” And in heeding the calls for revolutionary change, the Next Generation Air Transportation System (NextGen) initiative was enacted to address the needs of an increasingly stressed system.

Advanced Airspace Concept (AAC)

Since the birth of the NextGen initiative, a number of concepts have been developed that serve as the fuel for the changes necessary for the air transportation system to be able to handle the forecasted increases in traffic and the issues that it would entail. Among these concepts is the Advanced Airspace Concept (AAC) developed by Dr. Heinz Erzberger at the NASA Ames Research Center (Erzberger, 2001; Erzberger, 2004). This concept involves changes to the existing ground-based equipment and cockpit avionics as well as the communication technologies that would facilitate the interaction between the two. Some researchers contend that current air space capacity limits are dictated by the workload that handling traffic imposes on air traffic controllers. It is envisioned that the changes involved in the AAC, along with the revised roles and responsibilities of the system players that these changes would enable, would ultimately allow for the predicted increases in air space capacity due to the beneficial effects on air traffic controller workload.

One of the primary contributors to an air traffic controller's workload lies in their responsibility for providing separation assurance for all aircraft. This refers to the responsibility that the controller has for ensuring that all aircraft under their control are safely separated by the prescribed separation minima of five nm laterally and 1000 to 2000 ft vertically depending on the aircraft's altitude. Because of its centrality both to safety and capacity, the area of separation assurance has been one of intense focus for a number of researchers. One particularly important area of research has been on the potential benefits that the development and use of automation would have in support of separation assurance.

A central component of the Advanced Airspace Concept is the further introduction of automation into the current suite of tools used both on the ground and in the air. In terms of separation assurance, this would involve the development of a ground-based computer system that is capable of automatically generating both efficient and conflict-free clearances in response to aircraft pairs that appear to be in conflict. These automatically generated clearances could then be executed by the aircraft. This latter step highlights another key element of the overall concept in that the clearance would be able to be transferred from the ground system directly to the Flight Management System (FMS) of the appropriate aircraft via Controller Pilot Data Link Communications (CPDLC) channels (Bolczak, Gonda, Saumsiegle, & Tornese, 2004). This would enable the clearances generated by the ground-based system to be uplinked to the aircraft without any required voice communication from the air traffic controller.

Within the context of the AAC and, more specifically, the ground-based computer system, the particular area that this study focused on was the automated conflict resolution algorithm that is responsible for generating the clearances outlined earlier. This algorithm, also developed by Dr. Heinz Erzberger, was designed to provide automated separation assurance through the analysis of a given conflict situation and a subsequent generation of a resolution. The main benefit that the use of this algorithm is said to provide is that, assuming the infrastructure necessary for the AAC to be realized is in place, conflict resolutions can be generated and automatically sent to the involved aircraft without the need for controller involvement. Such advancements would, presumably, offload a great deal of the air traffic controller's workload to the automation, which would then allow for an increased capacity of the airspace that they are responsible for.

While there have been other tools and algorithms developed for the conflict resolution task such as MITRE's Problem Analysis, Resolution and Ranking (PARR) enhancement (Kirk, Bowen, Heagy, Rozen, & Viets, 2001) and Eurocontrol's Conflict Resolution Assistant (CORA) (Kirwan & Flynn, 2001) (see Kirwan & Flynn, 2002 for an extensive literature review of existing conflict resolution tools), the one used in this study is unique in that it not only accounts for a wider range of conflict types than the others but it also accounts for the various ways in which conflicts can be and are, in fact, resolved by air traffic controllers. Perhaps a more important point of uniqueness relates to the role that the automated algorithm is ultimately envisioned to assume. While most of the other conflict resolution tools have been developed to assume a supporting role, the tool in this study is proposed to occupy a niche within the air traffic management architecture that would remove the human from the conflict resolution equation thus permitting it to orchestrate the resolution of conflicts autonomously. As this is a rather radical proposal, the functioning of the conflict resolution tool/algorithm warrants an investigation into a number of relevant issues.

Before detailing the issues that were investigated in this study, however, a brief description of the algorithm and its logic will be presented. A more detailed description of the algorithm can be found in Dr. Heinz Erzberger's paper, "Automated Conflict Resolution for Air Traffic Control" (2006).

Automated Conflict Resolution Algorithm

The conflict resolution algorithm is basically made up of four components, arranged hierarchically, which operate iteratively in both a feedforward and feedback manner (see Figure 1). The algorithm takes as input information associated with a pair of aircraft (e.g. altitude, speed, coordinates) that are predicted to lose separation. This information is passed to the Resolution Aircraft and Maneuver Selector (RAMS), which identifies the type of conflict under consideration (e.g. both aircraft at level flight, one aircraft climbing, etc.) and generates a prioritized set of possible resolution maneuvers accordingly. This prioritization is based on RAMS's identification and selection of the aircraft most appropriate for the maneuver in addition to the type of maneuver that is to be executed. Once this set has been constructed, it is then fed to the Resolution Maneuver Generator (RMG). Based on the set of resolutions generated by the RAMS, the RMG creates simplified trajectories for those resolutions. The first maneuver template in the prioritized list from the RMG is then fed into the 4D Trajectory Synthesizer (TS). It is at this point that a finalized version of a flyable trajectory is formulated. This final formulation includes and accounts for more detailed information concerning a number of variables such as atmospheric conditions, aircraft performance, and operational procedures. Before continuing, however, the trajectory is verified to be indeed flyable by the selected aircraft. In the event that it is not, a message is sent back to the RMG for the next maneuver in the set to be sent forward for final trajectory

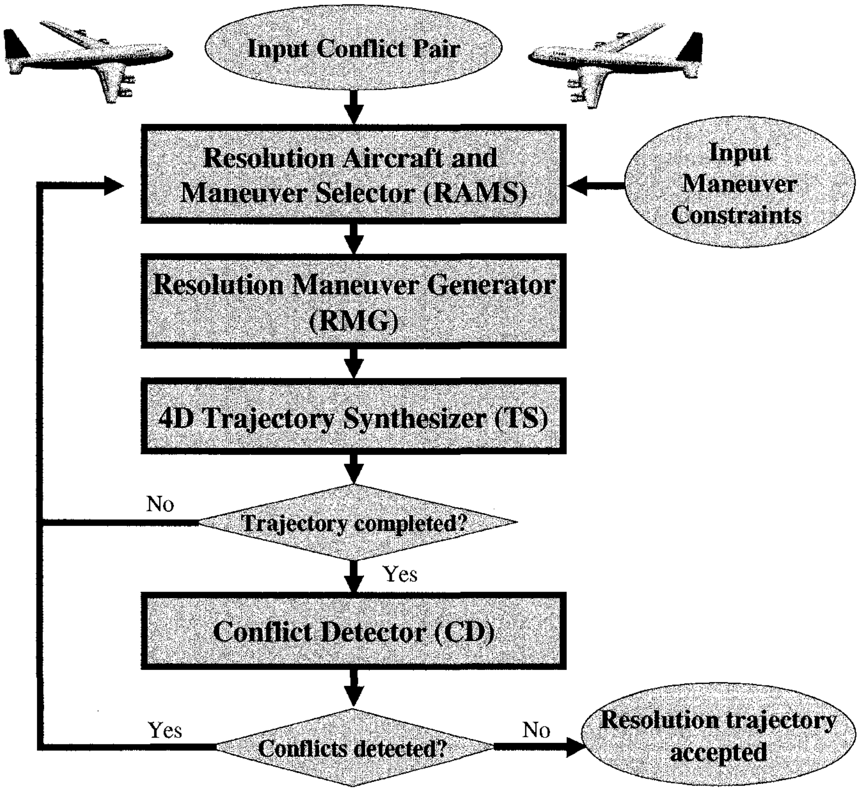


Figure 1. Flow chart of the resolution algorithm.

Note. From "Automated conflict resolution for air traffic control," by H. Erzberger, 2006, *Proceedings of the 25th international congress of the aeronautical sciences*, p. 2. Copyright 2006 by H. Erzberger. Reprinted with permission.

formulation and verification. Once an acceptable trajectory is formulated by the TS, it is sent to the Conflict Detector (CD). Here the resolution trajectory is checked in order to ensure that it will successfully resolve the current conflict, and that in doing so it has not inadvertently created a secondary conflict downstream. Similar to the communication between the TS and RMG, in the event that the finalized resolution trajectory does not successfully resolve the conflict, a message is sent back to the RMG with a request for the next resolution in the list to be sent forward for review. Each time a failure like this occurs, the algorithm uses that case as diagnostic information for consideration in the formulation of subsequent trial resolutions. Once a finalized, conflict-free resolution is generated, it is finally ready for transmission to the intended aircraft.

Having outlined the manner in which the algorithm generates conflict resolutions, it is important to explain some of the logic that underlies their formulation. Built into the algorithm is the ability to set the priorities and order for the types of resolutions that would be tried for a given conflict. For example, in the current study the algorithm was implemented as a horizontal first resolver. This meant that the process of generating a conflict resolution followed a preferential order such that lateral maneuvers were attempted first before moving on to altitude changes. It should be noted that the algorithm's design does include an additional capability of generating speed resolutions but that it was not implemented in the current study. In terms of lateral resolutions, the way in which they are implemented is through the positioning of a waypoint that is offset from the aircraft's original flight path a sufficient number of degrees to clear the conflict. An additional waypoint serves as the position where the maneuvering aircraft would rejoin its original flight path following its conflict avoidance maneuver. A guiding principle in the selection of these waypoints is the minimization of flight path deviation and delay.

In the event that the magnitude of flight path deviation becomes too extreme or a flyable solution simply does not exist, the algorithm next attempts to resolve the conflict through a change in altitude. In this case, the preference is to first look for resolutions involving a climb as they are generally considered more fuel efficient. In this study the algorithm searched for conflict-free altitudes in 1000 foot increments, not accounting for the maneuvering aircraft's direction of flight. If a climbing resolution cannot be found, a descent is then attempted in the same manner. Similar to the way in which lateral resolution trajectories are flown, altitude changes involve the aircraft climbing or descending to the assigned altitude and temporarily maintaining that altitude for a specified period of time. The aircraft then returns to its previous altitude and rejoins its original flight path.

In addition to the factors that are weighed in the formulation of lateral and altitude resolutions, an overriding factor depends upon which of the aircraft in the conflict pair is chosen to be the maneuvering aircraft. One of the most deciding factors in this case is if one of the aircraft in the conflict pair is an arrival or departure. For example, in most cases, the overflight aircraft would be chosen as the maneuvering aircraft so as to not impede the progress of the other aircraft.

By virtue of the resolutions being generated by an algorithm that has set parameters, a proposed benefit of using it, in addition to a reduction in air traffic controller workload, is consistency. Use of the algorithm would arguably provide consistent benefits in terms of the efficiency of flight due to the fact that it has a defined strategy that is geared towards efficiency and that this strategy is applied in a repetitive and consistent manner. This may be in contrast to air traffic controllers, who, by virtue of being an aggregate of individual human beings, exercise different strategies in resolving aircraft conflicts.

It is as of yet unclear whether or not the consistency offered by the algorithm provides the same, if not better, results in terms of system performance than what human air traffic controllers would in the same environment. This is an important measure to address because, after all, any proposed piece of automation must first be able to provide at least equivalent levels of performance of assigned tasks as the human it would ultimately complement or replace. This requires a measure of comparison between the automated algorithm and the human in identical situations. Additionally, it would be important to make this comparison not only in the context of current day levels of traffic but also at the levels that the NAS is projected to reach in the near future.

Current Study

The purpose of this study was to establish this comparison between the human and the automation in terms of system performance with the intent of identifying how each agent performs both individually across progressively higher levels of traffic density as well as how they perform relative to one another. In doing so, it was hoped that their strengths and weaknesses would come to bear with respect to the ability to safely and efficiently manage the traffic as well as how that might change as the level of traffic increases. An additional intent of this study was to gauge the human air traffic controllers' acceptability of the manner in which the automation performs the task of conflict resolution and how access to that automation by the human impacts their workload.

Levels of Comparison

As the automated conflict resolution algorithm was designed for the task of resolving conflicts, the focus of comparison in this study was isolated to that task. Therefore, the measures of system performance had to be framed in terms of what impact the potentially different ways that the human and automation resolve aircraft conflicts had on those measures. Also, as the scope of this study was limited to conflict resolutions, the traditional issues of automation (e.g. bias, failure, situational awareness, etc.) were not examined. This also meant that, for comparison purposes, the responsibilities of the air traffic controller were strictly limited to conflict resolution.

As mentioned earlier, the measure of comparison between the human and automation was going to be in terms of system performance. For the purposes of this study, system performance was defined as the safety and efficiency of the traffic as it transited a prescribed sector. More specifically, safety was defined as the number of separation violations that occurred as well as the overall minimum separation distances of the aircraft over a certain amount of time. Efficiency was further defined as the time and lateral and vertical distance added to the airspace system as a result of the resolutions executed. Acceptability was defined by the air traffic controllers' subjective impressions of the automation as well as more objective measures related to the way in which they used the automation when it was made available.

With respect to the levels of comparison that were instituted in this study, the first basic level dealt with traffic density. As the increases in traffic density for the NAS are predicted to grow to up to three times current day levels, this study investigated conflict resolution performance at three separate traffic density levels: current day (1X), twice current day levels (2X), and three times current day levels (3X).

The other level of comparison was related to the conflict resolution mode. There were three separate modes investigated in this study: manual, interactive, and fully automated.

The title of manual mode of conflict resolution in this study is a bit of a misnomer as it implies that the air traffic controller has no automated tools available to assist in the task. However, in this study, the air traffic controllers operating in the manual mode had automated conflict detection and a trial planning function. The data gained from this aspect of the study served as both a point of comparison for the fully automated mode as well as an opportunity to observe the air traffic controllers of today attempt to cope with the potential traffic environment of tomorrow. Perhaps due to the uncertainties of what changes will be introduced in the future this is a surprisingly understudied area of research that is deserving of analysis in its own right.

The interactive mode involved the same basic environment as found in the manual mode with automated conflict detection and trial planning capabilities, but the key difference was that the air traffic controller now had access to conflict resolution suggestions provided by the automated algorithm. The suggestions offered by the algorithm were able to be accepted, modified, or rejected depending on the assessment by the controller. The inclusion of this interactive condition was a necessary component due to the understanding that any introduction of an automated tool for conflict resolution would likely need to be phased in first in a supporting role before being able to assume the full burden and responsibility for the conflict resolution task as is envisioned. Thus, the controllers' performance while having the automated tool available would need to be assessed in order to ensure that there was some benefit it terms of performance and workload relative to when the tool was not available. An additional benefit to including the interactive condition was that it provided an implicit and objective means by which the acceptability of the algorithm's suggested resolutions and ultimate functioning could be gauged.

The final resolution mode was the fully automated mode in which the algorithm resolved conflicts completely autonomously. Once again there was automated conflict detection from which the algorithm based its resolutions on. The fully automated mode not only served as an important point of comparison with the human air traffic controllers but provided the opportunity to observe and analyze the behavior and potential of the algorithm in the future operations for which it was designed for.

Research Questions

As this was the first human-in-the-loop study of the Advanced Airspace Concept's automated conflict resolution algorithm, its conduct was more exploratory in nature. Through the levels of comparison just outlined, a number of questions were addressed.

Given that the chief purpose of the automated algorithm is to provide separation assurance, the first question addressed was what kind of impact the different resolution modes had on the system performance metric of safety across the three traffic density levels? This involved an analysis of separation violations and minimum separation distances.

The second question that this study addressed was how the efficiency of flights was affected by the different resolution modes, and how they might have changed as a result of the increasing traffic density levels. The answers to these questions resided in the system performance metric of efficiency with respect to the total delay added to the system as a result of the resolutions as well as the lateral distance added and vertical distances traversed. With its relationship to time and expense, the topic of efficiency was assessed as it ranks high on the list of airspace system users' priorities.

The third question investigated concerned how the increases in traffic levels and, more importantly, access to the automated algorithm in support of the conflict resolution task affected workload relative to when the automation was not available. This was an important measure to address as one of the primary benefits of the algorithm is to reduce the air traffic controller's workload. Although this benefit refers more to the stage at which the algorithm is functioning autonomously, it would still need to be examined in the context of a transitional, supporting role that it would most likely need to fulfill first.

The final question that this study addressed related to the acceptability of the algorithm's resolutions. More specifically, what is the general acceptability of the resolutions generated by the algorithm both in terms of subjective impression and willingness to actually implement the resolution at each level of traffic density, and does acceptability change as the levels increase? The issue of acceptability was necessary to include as the introduction of any tool into the air traffic controller's arsenal would face stiff opposition and potentially impassable roadblocks without it.

METHOD

Participants

Nine participants volunteered to take part in this study. Four of the participants were retired air traffic controllers with over 20 years experience. Four of the other participants were general aviation pilots and recent graduates of San Jose State University's Aviation program. One participant was a student preparing to enter the air traffic controller training and selection program. Participation was solicited via email with subsequent inclusion based on availability. During the planning and recruitment process, participants were assembled into groups of three as the intent was to run three participants in parallel. With a total of nine participants, this meant that there were three groups. As this was a within-subjects study the composition of each group was not a concern with respect to bias. All participants were paid for their time. There were also nine supporting participants that were all general aviation pilots.

Apparatus and Stimuli

Participants used individual personal computer (PC) workstations equipped with the Multi-Aircraft Control System (MACS) software (Prevôt, 2002) and connected to a 28" display monitor. The equipment that the participants used was located in the Airspace Operations Laboratory (AOL) at the NASA Ames Research Center and can be seen in Figure 2. Supporting pseudopilot functions were performed with networked computers from an adjacent room in an automated mode. These stations were monitored by supporting participants for situations in which manual intervention was needed.

Figure 2. MACS workstations in the Airspace Operations Laboratory.

Objective data was logged at each participant and supporting workstation as well as at dedicated data collection stations through MACS's data collection processes. Movie screen captures of each run were recorded with the commercial Camtasia Studio software at each participant's workstation. Subjective data on workload was gathered at five minute intervals during each run via MACS's workload assessment keypad emulation. Subjective data was further gathered through the presentation of questionnaires presented to the participants at the conclusion of each run (see Appendix A).

The objective data collected from the participants were in reference to their performance in resolving conflicts at three progressive levels of traffic density and two different conflict resolution modes. This was made possible by MACS and the tailoring of the system by Dr. Thomas Prevôt and his software development team for the needs of this study.

One major development in MACS that was introduced for this study concerned the display system replacement (DSR) graphical emulation that the participants used for controlling air traffic. One configuration of the DSR display in MACS provides the user with a nearly identical display to what is seen in today's en route ATC environment. However, with the increased traffic levels that were planned for the study, it was clear that keeping the display in its current day configuration would not be a viable option. For example, as shown in Figure 3, aircraft present on a current day DSR display are each represented by an aircraft position symbol with an associated data block that contains basic information about that aircraft such as its callsign, assigned altitude, and speed.

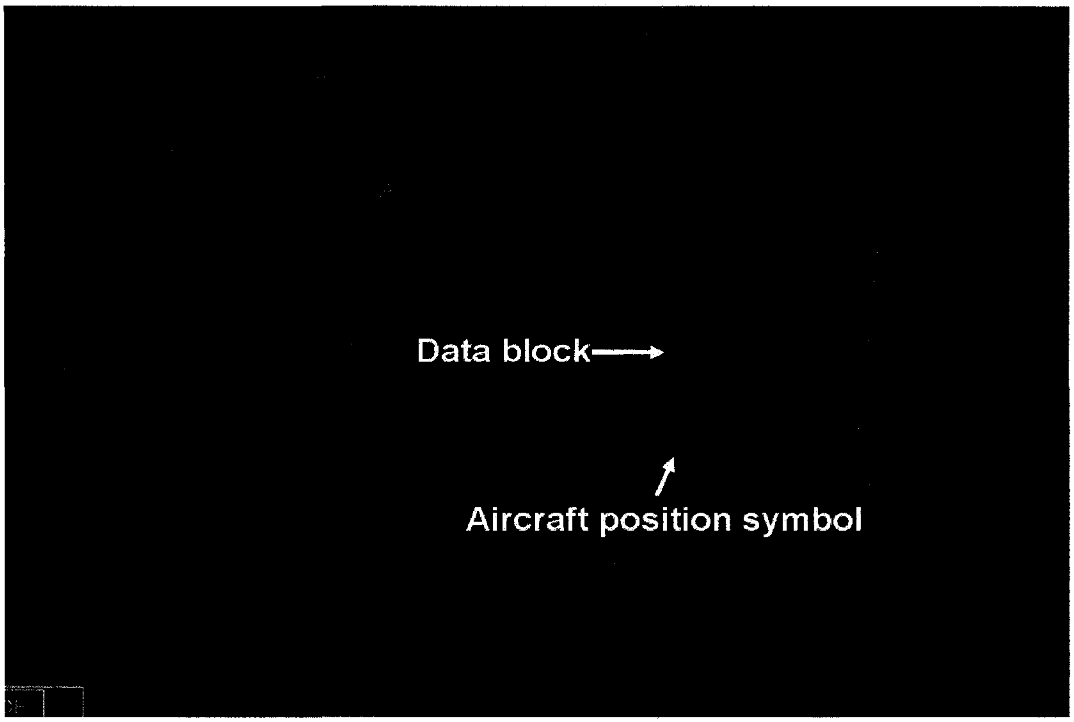


Figure 3. Current day display in MACS with basic information highlighted.

By keeping this setup, particularly at the 3X level of traffic, the display became so cluttered with all of the aircraft's data blocks (see Figure 4) that an individual would have to devote nearly all of their time and effort to decluttering the display with little time left to resolve conflicts. Even if one were to forego decluttering the display and simply attempt to resolve conflicts using the current day configuration, the clutter from all of the symbology and data blocks would become so unruly that resolving most conflicts would be nearly impossible.

These difficulties led to the modification of the current display to one in which aircraft that were not in conflict had collapsed data blocks and the color of the symbology for these aircraft were dimmed down to a point where they nearly blended in with the background (see Figure 5). This modification was also in keeping with one of the purposes of this study, which was to isolate and limit the tasks of the controller to conflict resolution, freeing them from the responsibility for attending to any of the other more routine tasks that they are currently responsible for.

In contrast to the dimmed down aircraft not in conflict, those that were in conflict were made salient to the participants by highlighting the aircraft symbols with different hues that indicated the time to imminent loss of separation (LOS). As shown in Figure 5, there were three colors used for this highlighting, with the aircraft in white representing a loss of separation in anywhere from 12 to eight minutes. Aircraft with yellow symbology represented a predicted LOS in between eight and five minutes. Finally, aircraft that were presented in orange represented a predicted LOS in between five and two minutes.

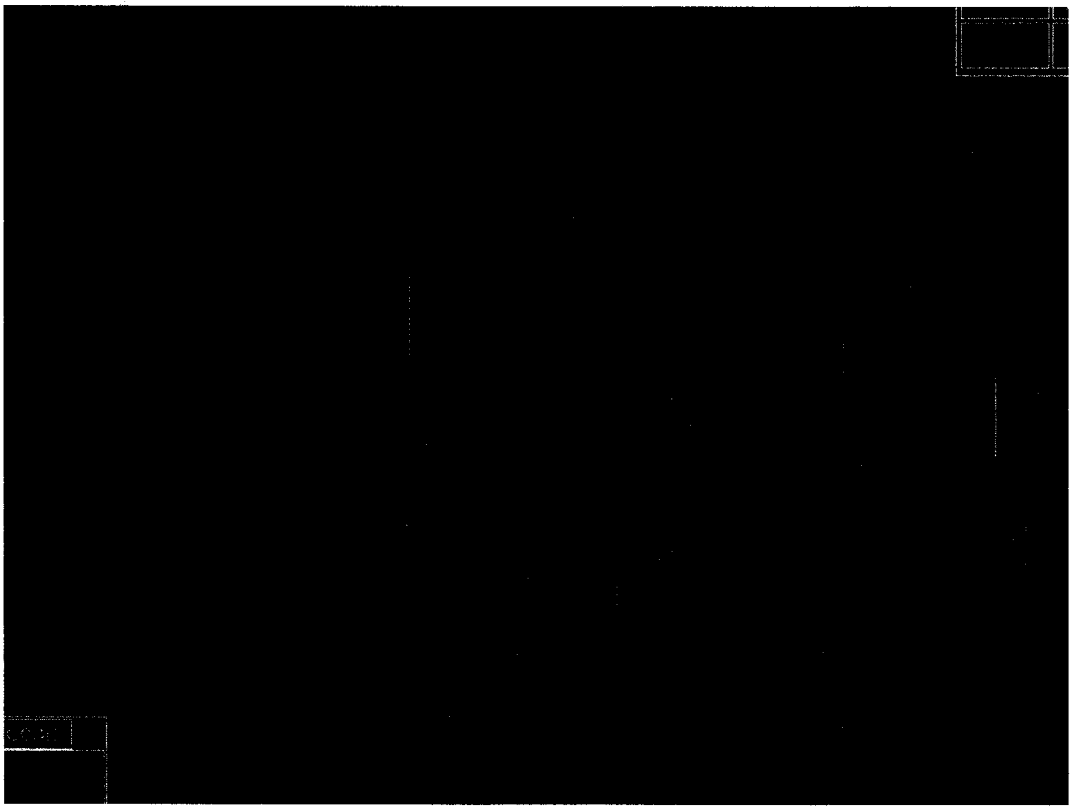


Figure 4. Current day display at 3X level of traffic.

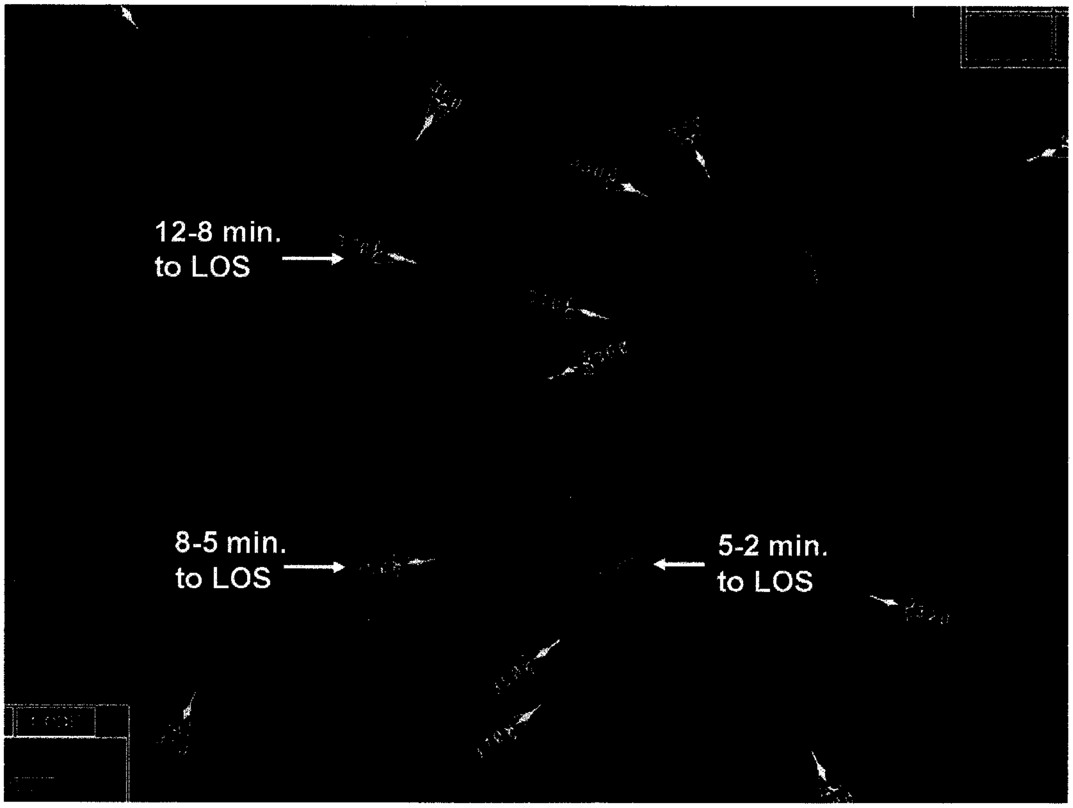


Figure 5. Advanced display settings at 3X level of traffic.

All of the aircraft in conflict were also shown to the participants in a conflict list, which had the conflict pairs ordered according to the times to loss of separation. This meant that those aircraft pairs that were predicted to lose separation most immediately appeared at the top of the list while those with greater times to LOS appeared further down the list.

With less than two minutes, the aircraft in conflict dropped out of the conflict list and became dimmed out once again. Additionally, their collapsed data tag flashed red until the aircraft pair was outside of the prescribed separation minima. This essentially ended the time available for resolving the conflict at which point the participant would move on to the next conflict. The reason for taking these measures was that the automated algorithm stopped generating conflict solutions with less than two minutes to LOS. Allowing the conflicting aircraft to remain highlighted past this two minute point would not have allowed for a fair comparison between the automation conditions as well as a fair assessment of how the tool was used when it was available. It should be noted here that the presentation of aircraft in conflict only applied to those that were predicted to lose separation within the boundaries of the test sectors. Those aircraft that would lose separation outside of the test sector were not presented to the participants as they were only responsible for conflicts happening inside of their sector.

With regard to simulated air traffic, it was decided early on to base the traffic scenarios on those used in a previous study that looked at Trajectory Oriented Operations with Limited Delegation (TOOWiLD) (Prevôt et al., 2007). The scenarios from this previous study were based on airspace that included sectors from the Kansas City Air

Route Traffic Control Center (ZKC) and Indianapolis Air Route Traffic Control Center (ZID) (see Figure 6).

A total of six scenarios were developed with two scenarios per level of traffic: 1X, 2X, and 3X. For the scenarios used in the current day (1X) traffic level condition, the same two scenarios used in the TOOWiLD study's current day, datalink condition were adapted. The only difference was that as a result of the scenarios from the previous study being 75 minutes in length with gradual traffic build-up times, a recording of both scenarios needed to be made that would tailor them to the target 30 minute duration time of this study. This was done with the MACS recording function by playing the original scenarios for approximately 20 minutes, which was the point at which the original scenarios had reached the peak traffic levels of approximately 15 aircraft in both ZKC 50 and ZID 91 test sectors. After reaching this point the recordings started. The end product was basically two separate 1X scenarios that mirrored the original scenarios except for the fact that they had, from the start, a fairly steady stream of approximately 30 aircraft at any given time transiting the combined test sectors shown in Figure 6. The two scenarios created here served as the basis for the remaining four scenarios at the 2X and 3X levels collectively.

For the creation of the two scenarios used to represent the 2X traffic level, the same two scenarios created for the 1X condition were used as their foundation. However, these 1X scenarios were altered in a number of ways in order to achieve the desired sector counts. As previously mentioned, the original scenarios that the 1X scenarios were based upon had a duration that was much longer than what was necessary for this study.

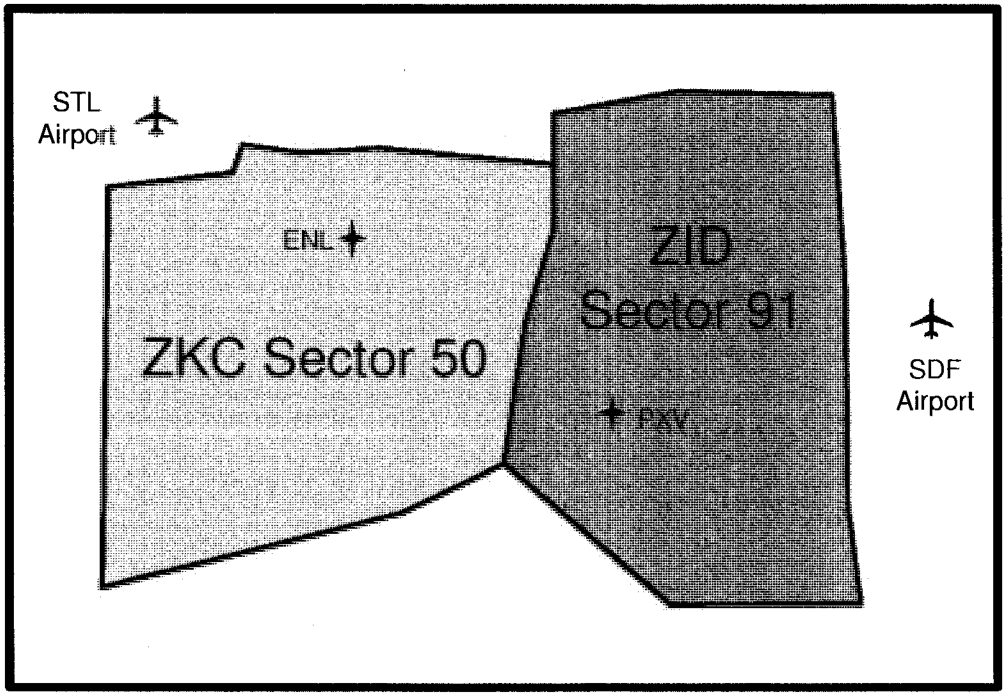


Figure 6. Airspace used in the study.

As a result, a number of aircraft from the 1X scenarios never entered the test sectors during the desired 30 minute time frame. Therefore, to achieve and maintain the necessary traffic levels of 60 aircraft in the 2X condition, those aircraft that were too far away in the 1X scenarios were moved closer to the test sectors while ensuring that they remained on their original routes. However, this left at least half of the aircraft in the scenarios identical to what was used in the 1X scenarios. Consequently, in order to avoid recognition of the aircraft by the participants, a number of steps were taken. One such step was to essentially randomize the callsign of each aircraft in the scenario to at least render the aircraft unidentifiable to the participant at the name level. A further step was taken to reduce the similarities of the 1X and 2X scenarios by randomizing the starting points and altitudes of all aircraft in both 2X scenarios. Care was taken to ensure that the new altitudes were consistent with their direction of flight such that aircraft flying in an easterly direction were at odd altitudes and those flying in a westerly direction were at even altitudes.

While moving aircraft closer to the test sectors was sufficient for generating scenarios at the 2X level, additional measures were needed in order to reach the desired traffic count of 90 aircraft at any given time in the combined test sectors at the 3X level. As each of the two scenarios in the 2X condition was based on one of the two 1X scenarios, the first step in creating a 3X level scenario was to take one of the 2X level scenarios and combine it with the 1X scenario that it was not based on. This was done by copying aircraft from the 1X scenario and inserting them into the 2X scenario until the proper traffic count was finally attained. Upon reaching this point for both of the final

3X scenarios, the same randomization process used in the development of the 2X level scenarios was applied.

Before proceeding to the details of how the number of conflicts in each of the scenarios was achieved, it may be important to explain an assumption that was made that guided the initial scenario development process. Although the aircraft in the 1X scenarios all followed structured routing, it became clear during the development process that this would have to be changed somewhat as the traffic levels increased. While every effort was made to maintain each aircraft's original structured routing for the 2X and 3X scenarios, it quickly became evident that this would not be completely possible due to the fact that as the traffic level increased, the areas in the test sectors where many of the transiting aircraft shared a common waypoint –ENL and PXV specifically (see Figure 6) - became so overwhelmed with traffic that the conflicts occurring in those areas became completely unmanageable because there was no room to maneuver. As a result, a number of aircraft transiting those high density areas were given routes that were offset a given distance and ran parallel to the other traffic so as to at least maintain that particular flow of traffic.

Having reached the desired traffic counts for all six scenarios with the route structures just described, the issue of conflict counts was addressed. In terms of the 1X scenarios, because they were considered representative of current-day levels of traffic, the number of conflicts that ultimately resulted in a separation violation from simply allowing the scenarios to run open-loop was established as the baseline number from which those in the 2X and 3X scenarios would be derived. After playing out the two 1X scenarios, the total number of conflicts in each were eight and six respectively. As an underlying goal in this scenario generation process was to provide a challenging traffic environment that would provide ample opportunities for the excitement of differences between conditions, the greater number of eight was originally chosen as the base number that the conflict counts in the subsequent scenarios would be based upon.

In determining how this initial number of eight conflicts could be scaled to meet the most appropriate number for the 2X and 3X scenarios, the relevant literature described the increase in conflict counts in relation to traffic levels as being non-linear. For example, Georges Maignan, in his 1992 article "Safety Aspects of Increased Capacity of Airport and ATS," referred to a formula proposed much earlier by Robert Machol (1979) from which Maignan based his conclusion that two and threefold increases in air traffic would result in four and ninefold increases in conflicts respectively. This conclusion has since been supported particularly with respect to aircraft with structured routing as seen in today's environment and, for the most part, the scenarios used in this study (Jardin, 2004). This rationale has also been adopted and implemented in subsequent studies on issues related to traffic density increases (Hoekstra, Ruigrok, & van Gent, 2000). As a result, the number of conflicts that were used for the 2X and 3X scenarios in this study was based on squaring the magnitude of traffic density increase and then multiplying that number with the original base number of eight. This meant that the 2X scenarios eventually had approximately four times the number of conflicts as the 1X scenario, which translated to approximately 32 conflicts. The 3X scenarios had nine times the number of conflicts as the 1X scenario, which were approximately 72 conflicts for the 30 minute time frame.

With the number of conflicts for each scenario decided, additional work was needed in reaching that level since there were too few in the scenarios at this point. The goal in generating these additional conflicts was that the scenario would have a fairly steady number of conflicts throughout its duration and that there would be a variety of conflict situations (e.g. arrival vs. cruise, cruise vs. departure, etc.). This required the scripting of conflicts through an iterative process in which each scenario was played out in real time and potential conflict pairings were identified. Many of the conflicts that were ultimately scripted were done so by adjusting the altitude of one of the aircraft such that the resulting trajectory would lead to a loss of separation without some form of intervention.

Although the desired number of conflicts was eventually reached, a few late changes altered these numbers somewhat. One of these changes was a result of the scenarios being developed using a wind environment that was different from what was planned for use in the final study. Changing the wind characteristics of the study's airspace resulted in a change to the nature of some of the conflicts such that the original counts were later found to be different. Another change involved the "smoothing" of aircraft routes due to the observation that some of the existing routes in the scenario contained rather unrealistically sharp kinks. It was found that if an aircraft was given a direct-to by deleting one of the kink's waypoints, it would result in distance and time savings that were out of the norm and unlike anything that would be seen today. As a result, many of the routes in the scenarios were smoothed out by strategically inserting waypoints at particular points along the route in order to reduce the angle between some of the existing waypoints. Great care was taken to allow all of the aircraft's routes through the test sectors to remain identical, but a few alterations were needed which also resulted in a minor change to the number of conflicts. Attempts were made to reclaim the previous number of conflicts with moderate success in the 1X and 2X scenarios, but one of the final 3X scenarios ultimately had 61 conflicts as opposed to its previous count of 72. However, this change in number did not appear to have detrimental effects as it proved to be a number that was just as difficult for the participants to deal with as the other 3X scenario, which contained 73.

In addition to the six test scenarios that were created, six training scenarios were developed through much the same process as described earlier. The main difference in this case was that instead of increasing the number of conflicts in the scenarios, the number was reduced in order to provide the participants with the opportunity to familiarize themselves with the flow of traffic through the test sectors. This reduced number also gave participants the time to get comfortable with using the display and tools for conflict resolution, which also allowed them to focus on the task itself without the need to worry about keeping up with the conflict frequency that they would encounter later during the test runs.

Before continuing on to the explanation of this study's experimental conditions, it may be important to explain some of the basic characteristics and assumptions that governed the air traffic environment that the participants operated in. An important assumption that was made was that the conflict detection system that alerted the participants to the conflicts was one hundred percent accurate and participants were informed that it should be trusted completely. This absolved the participant from having to monitor and search for potential conflicts and also served to limit the scope of this experiment solely to the task of conflict resolution. With respect to the actual aircraft in all of the scenarios, all were data link equipped and automatic dependent surveillance – broadcast (ADS-B) equipped with a one second update rate on the display. As far as control and ownership of these aircraft was concerned, the participant owned all aircraft from the start of the scenario and could send clearances via data link to any aircraft regardless of whether or not it was in the test sectors. Additionally, the participants were only alerted to and responsible for conflicts that would result in LOS within the test sectors. The participants were not alerted to nor were they to be concerned with conflicts that would happen outside of the test sectors.

Conflict Resolution Conditions

The purpose of this study was to investigate the performance of humans and automation in the task of conflict resolution across different levels of traffic density. This resulted in a total of nine experimental conditions:

1. **1X Manual:** Participants managed approximately 30 aircraft at any given time with the aid of an automated trial planning tool for the construction of resolutions (see Figure 7).
2. **2X Manual:** Participants managed approximately 60 aircraft at any given time with the aid of an automated trial planning tool.

3. **3X Manual:** Participants managed approximately 90 aircraft at any given time with the aid of an automated trial planning tool.
4. **1X Interactive:** Participants managed approximately 30 aircraft at any given time with the aid of an automated trial planning tool. Additionally, the participants were equipped with the conflict resolution algorithm, which automatically generated and made available conflict resolutions once activated.
5. **2X Interactive:** Participants managed approximately 60 aircraft at any given time with the aid of an automated trial planning tool and the conflict resolution algorithm.
6. **3X Interactive:** Participants managed approximately 90 aircraft at any given time with the aid of an automated trial planning tool as well as the conflict resolution algorithm.
7. **1X Fully Automated:** The automated conflict resolution algorithm attempted to resolve all conflicts without human involvement in an environment with approximately 30 aircraft occupying the test sectors at any given time.
8. **2X Fully Automated:** The automated conflict resolution algorithm attempted to resolve all conflicts without human involvement. Approximately 60 aircraft occupied the test sectors at any given time.
9. **3X Fully Automated:** The automated conflict resolution algorithm attempted to resolve all conflicts without human involvement. Approximately 90 aircraft occupied the test sectors at any given time.

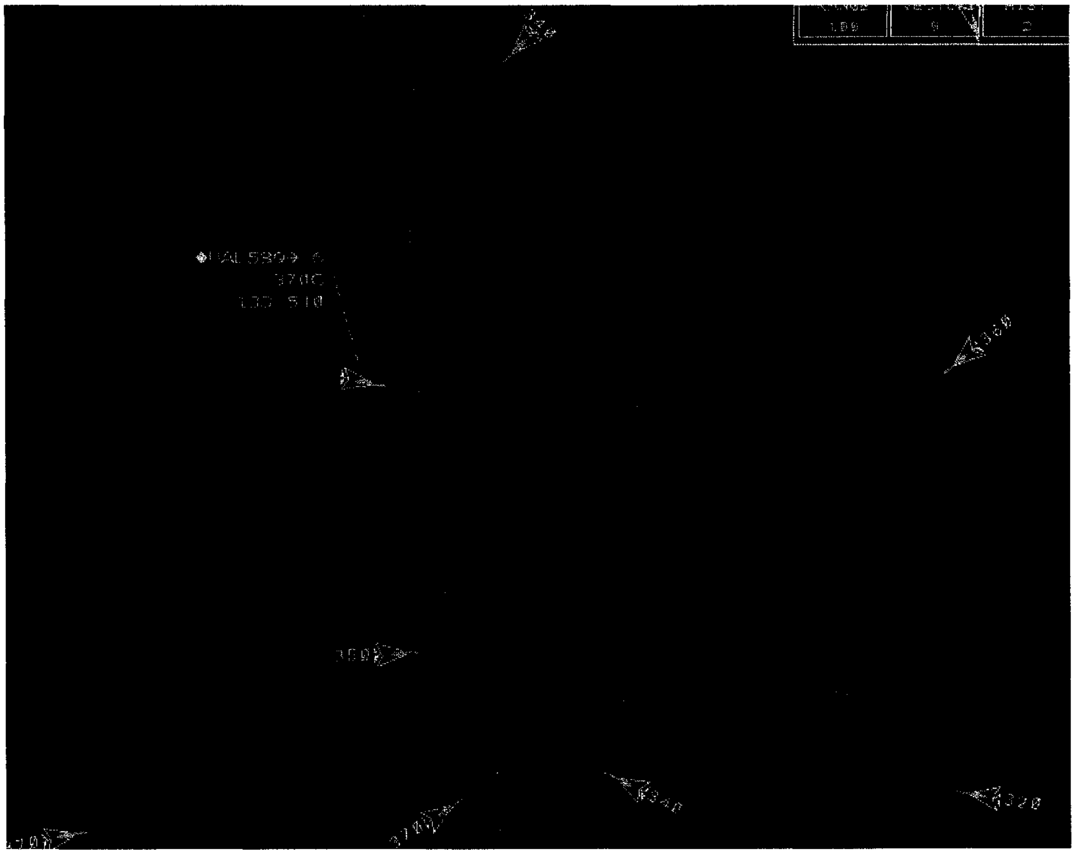


Figure 7. Example of the manual trial planning tool used for conflict resolution.

Design and Procedure

In this 3x3 within-subjects repeated measures design, the two independent variables (IV) that were manipulated across conditions were traffic density and conflict resolution mode. The traffic density IV consisted of three levels: 1X, 2X, and 3X. The conflict resolution mode IV consisted of three levels: manual, interactive, and fully automated. The matrix created by the mixture of these IVs and their associated levels consisted of nine conditions as outlined above. However, as the fully automated conditions did not involve any human intervention, participants were only exposed to the six conditions that related to the manual and interactive modes of conflict resolution.

The dependent variables that were used to measure the effects of these independent variables dealt with system performance metrics related to safety, efficiency, workload, and acceptability. Safety was measured in terms of separation violations and minimum separation distances. Efficiency was measured by the total delay imposed on the system, the average delay per resolution, the average lateral distance deviation per lateral resolution clearance, and the average vertical distance traversed by aircraft given altitude clearances. Workload was measured during each run through prompts every five minutes on each participant's display. The peaks and means of this metric were included in the final analysis. An additional measure of workload was obtained through post-run questionnaires that, in part, asked participants to rate their peak workload for the run just completed. The final dependent measure was acceptability, which included separate objective and subjective components. The objective measurement was taken from the participants' usage characteristics of the automated conflict resolution algorithm when it was available for assistance. This involved how often the automation was used, and how often the suggested resolutions were accepted or rejected. The subjective measure of acceptability was taken from the responses that participants gave to the post-run questionnaire following runs in the interactive condition.

With the independent variables decided upon, the instruments in place, and receipt of the proper approvals (see Appendix B), the formal experiment began in the Airspace Operations Laboratory at the NASA Ames Research Center. With respect to the runs in which participants were involved, there were a total of 12 runs – one run per each of the six conditions with one repetition. All participants were exposed to all conditions with the order of runs following the schedule presented in Table 1. The administration of these runs was accomplished over the course of two days.

Upon arrival of the first group of participants on the first day of the experiment, they were given an initial briefing on the details of the study and what would be expected of them over the next two days. The participants were then asked to read the consent form (see Appendix C) and, if the terms and conditions were deemed agreeable, sign. Following this step, the participants were seated at one of the three workstations configured for the study, after which the training commenced. This involved first loading a 1X training scenario into each of the three MACS simulation manager stations –one for each participant- and then playing it. Each participant trained and later performed for the data collection runs in parallel such that all three workstations were always running the same exact scenarios with the same exact configuration. Additional preparation steps included the initiation of the supporting controller and pilot stations for each of the three

Table 1.

Study schedule used throughout study.

Day 1

9:00	Admin./Brief		
9:30	Begin training		
12:15	Lunch		
1:15	Resume training		
2:00	Data Collection: Run 1	1X	Manual
2:50	Run 2	2X	Interactive
3:40	Run 3	3X	Manual
4:30	Run 4	1X	Interactive

Day 2

9:00	Run 5	2X	Manual
10:00	Run6	3X	Interactive
11:00	Run 7	3X	Interactive
11:30	Lunch		
12:30	Run 8	2X	Manual
1:20	Run 9	1X	Interactive
2:10	Run 10	3X	Manual
3:10	Run 11	2X	Interactive
4:00	Run 12	1X	Manual
4:30	Debrief/Make-up runs if needed		

clusters associated with each participant in the rooms adjacent to the experimental room. At this point, however, the data collection stations were not activated. The stations within each of the three participant clusters were linked via the aeronautical datalink and radar simulator (ADRS) system.

With the lights dimmed in the experiment room and the first training scenario running, each participant was trained on the basics of the MACS DSR display with particular focus on the composition of the data blocks, aircraft symbology, data link operations, and the manual trial planning tool. They were also briefed and given the chance to familiarize themselves with the airspace that they were responsible for and the traffic flow characteristics of that airspace. Additionally, the participants were briefed on their responsibilities being limited to conflict resolution only and that the task was further limited to only resolving those conflicts that were presented to them as being predicted to occur in their sector. The two 1X training scenarios were run once, which gave the participants ample time to get a feel for the basic operations that they would be asked to perform in the experimental runs.

Following the training with the 1X scenarios, the 2X training scenarios were loaded at each simulation manager station and subsequently played. These scenarios had a greater number of conflicts at a greater yet still relaxed frequency. This gave the participants numerous chances to practice the different methods of resolving conflicts manually while getting a better feel for the tools and the simulated environment. The 2X scenarios were played approximately three times with the first run being devoted to resolving conflicts manually and the remaining two runs being devoted to resolving conflicts with the automated resolution tool. At this point, the participants were fairly well practiced with using the tools and sufficiently familiar with the simulated environment whereas they were prepared to tackle the more complex problem presented by the 3X environment. Therefore, the final run before lunch was an introductory 3X training scenario. Aside from the schedule of training runs, the participants were given 15 minute breaks at the bottom of each hour until the lunch break.

At the conclusion of the first 3X training scenario, participants were given a one hour lunch break after which they returned to run one final 3X scenario before commencing the data collection runs. Following this final training run, the participants were given a 15 minute break while the laboratory was prepared for data collection. This basically involved restarting all of the stations associated with each participant, loading the first scenario, and ensuring that the data collection stations were running and that each of the relevant stations were actively linked. Each of the three simulation manager stations had the start times for the run synchronized such that all three stations would start the air traffic at the same time automatically.

Prior to the start of the data collection runs, the participants were given a final briefing in which they were reminded of their roles and responsibilities. Additionally, because the experimental 3X scenarios were known to be quite difficult, particularly in the manual conditions, the participants were told that if at any time they felt uncomfortable or overly frustrated during any run they were free to withdraw from the run or the entire experiment if so desired without any adverse consequences. And in fact this did actually happen on four separate occasions all during a 3X manual run.

Following this quick brief, the movie captures on each workstation were initiated, the lights in the experiment room were dimmed, and the first of four data collection runs for the day began. This and all other runs were 30 minutes in length. During that 30 minute period, the participants attempted to resolve all of the conflicts that they could while rating their subjective workload every five minutes through the workload assessment keypad appearing on their display. Meanwhile, supporting participants in the adjacent room were monitoring the associated support stations to intervene with aircraft out of performance or to take care of datalink messages that had not been automatically accepted by the intended aircraft. It should be mentioned here that for the entire duration of the experiment, the supporting participants only very rarely needed to step in and perform any action. At the 30 minute mark, the lights were brought back up to full intensity and the participants were informed of the run's completion. They were then given a post-run questionnaire to fill out after which they were free to go on a 15 to 20 minute break. In the meantime, the scenarios were kept running for an additional 10 minutes to allow for the traffic to exit the test sectors. This meant that each scenario was run for a total of 40 minutes although the participants were only required to resolve conflicts for the first 30.

After this 10 minute extension, all of the stations in each of the three clusters were brought down and then restarted. The next scenario was then loaded into the simulation manager stations with the timers set to start the traffic at the next scheduled time. At this point, the participants had returned from their break and had assumed their place at the same workstation as before. The movie captures at each station were then started and the lights were dimmed in preparation for the start of the next run. At the preset time, the traffic in the next scenario started and the participants performed the same conflict resolution task as before for the next 30 minutes.

This cycle of procedures just described was performed for all four data collection runs on the first day. At the conclusion of the final run, after filling out the questionnaire, the participants were reminded of the next day's start time and were then free to leave. At this point, all of the data collected from each of the participants' runs were compiled, organized, and stored on password protected computers. There was also no immediately identifiable information associated with any of the participants on any of the files. All of the consent forms and questionnaires were also locked in a secure room for storage.

As shown in Table 1, the schedule for the second day consisted of a total of eight data collection runs. The runs were divided evenly with four in the morning and four scheduled for the afternoon. The administration of these runs was much the same as how it was accomplished on the previous day. The only difference was that due to limited space available for data storage on some of the computers, the lunch break was used as an opportunity to clear out the data from the morning's runs in preparation for its eventual organization and storage at the end of the day.

At the conclusion of all 12 runs, the participants were debriefed and thanked for their valuable contributions. This session also provided an opportunity for an open discussion in which a variety of issues and concerns ranging anywhere from the overall operating concept to the behavior of the algorithm could be addressed. These debrief sessions provided valuable insight and informed some of the planned changes to the algorithm and MACS software following the completion of this study. Following the debrief, the data collected during the day was compiled, organized, and added to the data collected from the previous day's four runs such that one virtual folder contained all of the data associated with all 12 runs for each of the three participants in the first group. Again, this data was stored on password protected computers in a secure room.

The steps just described were repeated for the remaining two groups of participants without deviation. Careful attention was paid to accurately and promptly organize and store all data collected from each run while simultaneously making every effort to protect and maintain the anonymity of all participants.

Having completed the data collection for all three groups, the final step was to collect the data for the fully automated condition. This was done over the course of one morning in two sessions. The first session involved initializing all three computer clusters much in the same way as when participants were involved, starting the movie captures at each station, and loading the first set of scenarios such that each of the three clusters would be working one of the three 1X, 2X, and 3X scenarios. After the start, the automated algorithm worked the traffic independently by solving all conflicts automatically without human assistance in the task. Staff members were present for these runs, however, to monitor the progress as well as to intervene in the event of a malfunction of sorts. No conflicts were resolved by any of the staff members; the automated algorithm was the only agent to close the loop in resolving conflicts for these runs. One parameter that the automated algorithm was working under that differed from what the participants were was that it would only attempt conflict resolutions on conflicts with eight minutes or less to LOS. In contrast, participants were allowed to solve conflicts from 12 minutes out. This differential was the result of the algorithm's creator's assertion that conflicts with greater than eight minutes to LOS stand a greater chance of turning out to be false conflicts and attempting to resolve them could lead to unnecessary clearances being sent to and flown by the aircraft.

After the completion of this first round of data collection for the fully automated condition, the workstations were all restarted and prepared for the final round much in the same way as for the previous round. In this case, the second set of scenarios was loaded and ran once the movie captures had been started.

Following this final set of data collection runs, the data from all of the fully automated runs were compiled, organized, and placed in the same location as all of the data collected from the participants. This meant that data from the entire study was organized into one central folder. However, due to the risk of loss, this folder was copied and placed in four separate locations, all with security measures in place to prevent unauthorized access.

RESULTS

This study sought to analyze the conflict resolution capabilities of humans and automation in terms of system performance across three progressively higher levels of traffic density: 1X, 2X, and 3X. System performance was broken down into the specific areas of safety, efficiency, workload, and acceptability.

It should be noted here that due to the assumption that the performance of the automated algorithm in the fully automated conditions would always yield identical results, they were only run once for a total of six runs- each of the three traffic levels had two associated scenarios. This was in contrast to the manual and interactive conditions with humans in the loop that were run a total of 12 times for all nine participants. Because of the homogenization of variance in the fully automated conditions that resulted from the single run, the data gathered could not be included for comparison in the inferential statistical analysis. The descriptive statistics from these runs were included, however, as a means of comparison and for identifying trends.

Another important item of note is that although there were originally nine participants from whom data were collected for in this study, only eight could be used in much of the final analysis. The reason for this change was that, as stated earlier, participants were always given the option of discontinuing their participation in a run if they ever felt uncomfortable or overwhelmed. In this case, one of the retired air traffic controllers quit during both exposures to the 3X manual conflict resolution mode conditions for just those reasons. Although this fact was taken as a broader data point in and of itself, the result was completely missing data that basically excluded this individual's data from being able to be adequately compared with the other participants' data.

Safety

Since the ultimate measure of performance for both the human and the automation in both today's and the future's traffic environment is safety, it was analyzed first so as to set a framework for the viability of the automation as well as to gauge the performance of the humans in this most critical area.

Separation Violations

The first measure of safety that was analyzed was the number of separation violations. This dependent measure was composed of a raw count of the number of aircraft in a given run that violated the separation minima defined by the FAA. A separation violation was said to have occurred if two aircraft passed each other within a distance that was less than five nautical miles laterally and 1000 feet vertically.

As seen graphically in Figure 8, the descriptive statistics for this measure show that for the manual mode of conflict resolution, there was a large increase in the number of separation violations as the traffic levels increased. At the 1X, current day level, the average number of violations was, as should be expected, minimal (M= 0.06, SD= 0 .18). The raw numbers show that across all 16 runs for this 1X manual condition, there was one separation violation. As the traffic level increased from the 1X to the 2X level in the manual condition, there was an increase in the average number of separation violations to 1.81 (SD = 1.56). Although this was a considerable increase over the 1X condition, the greatest increase in the number of separation violations for all conditions in the study was

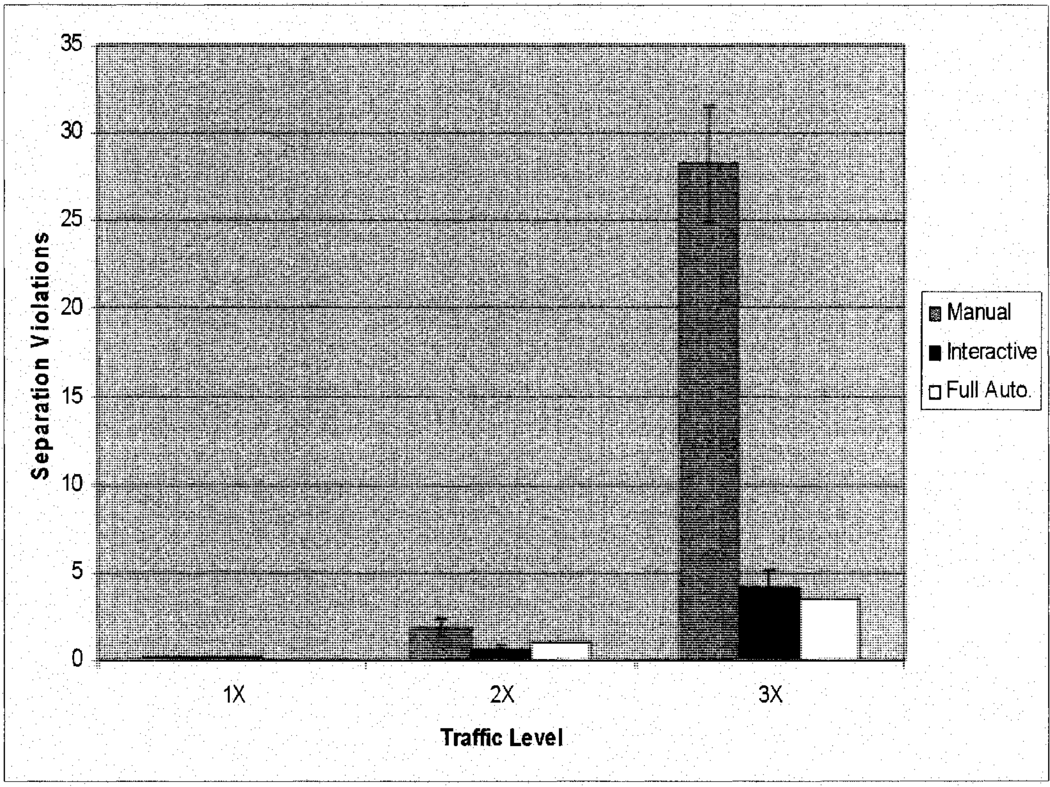


Figure 8. Separation violations for each conflict resolution mode and traffic level.

when the participants were exposed to the 3X level of traffic in the manual conflict resolution mode (*M*= 28.19, *SD* = 9.4). The resultant number of 28.19 translates to nearly one separation violation per minute of each trial in the 3X manual condition, which would be unacceptable by today's standards for safety.

In terms of the descriptive statistics for the interactive condition, one might notice immediately that the numbers are clearly different at the higher levels of traffic. Since the number of separation violations at the 1X manual level was so low, it was not surprising to see similar numbers in the interactive condition for the 1X level. In fact, the numbers were identical (*M*= 0.06, *SD*= 0.18). Again there was only one separation violation across all runs in this condition. The first departure, then, in terms of differences between the manual and interactive conditions came at the 2X level of traffic, which saw a reduction in the average number of separation violations (*M*= 0.56, *SD* = 0.56). A further departure between the manual and interactive conditions came at the 3X level of traffic. As before, there was an increase in the number of violations, but it was not as stark as in the manual condition (*M* = 4.13, *SD* = 2.80).

For the fully automated conditions, the descriptive statistics reveal performance characteristics that were somewhat more in line with those observed in the interactive conditions. At the 1X traffic level, there were no separation violations (M = 0.00, SD = 0.00). At the 2X level, however, the average number of separation violations saw an increase to 1.00 (SD = 0.00). There was a further increase in the number of separation violations at the 3X level of traffic, but the fully automated conflict resolution mode resulted in the fewest average number of violations out of all of the conflict resolution modes (M= 3.50 ,SD = 0.00). Note that the results for the fully automated condition all had standard deviations of 0.00. This was because the runs for this condition were not repeated due to the operating assumption that the automation would always behave identically in a given scenario regardless of the number of iterations. Therefore, there was nothing to compare the results from the single runs to in order to measure standard deviation. More conceptually, even if the trials in the fully automated condition were to be run multiple times, there would not be any standard deviation to report because of the consistency of the automation.

Based on these descriptive statistics, the use of the automated conflict resolution algorithm in both the interactive and fully automated modes had a clear advantage over the manual mode of conflict resolution in terms of separation violations. This advantage was seen most clearly at the 3X level of traffic where the manual mode conditions saw an average of 28.19 separation violations in contrast to the 4.13 and 3.50 of the interactive and fully automated modes respectively. However, in order to further investigate the statistical significance of these differences, a 2 (manual vs. interactive resolution mode) x 3 (1X, 2X, 3X traffic levels) repeated measures analysis of variance (ANOVA) was conducted on the number of separation violations. As a reminder, the fully automated conditions were not included in this stage of the analysis.

The inferential statistics for the analysis of separation violations revealed significant main effects for the mode of conflict resolution, $F(1, 7) = 72.13$, $p < .01$, as well as traffic level, $F(2, 14) = 62.30$, $p < .01$. There was also a significant interaction between the two variables, $F(2, 14) = 51.91$, $p < .01$, suggesting that there was a certain point at which the use of automation had a significant impact on the number of separation violations that occurred. Through subsequent paired *t*-tests of the manual and interactive resolution modes at each traffic level, it was found that the point at which this occurred was seen as early as the 2X level of traffic, $t(7)= 2.59, p< .05$, and continuing on to the 3X level of traffic, $t(7)= 7.69, p< .01$. These results point to the potential benefits that the automated tool could have in terms of safety in both the near and far term levels of traffic that are predicted to occur.

Minimum Separation Distances

In McNally and Gong's "Concept and Laboratory Analysis of Trajectory-Based Automation for Separation Assurance," (2006) they outline the minimum separation metric that was developed as a means of comparing the number of times that unique pairs of aircraft pass within or near the legal separation distances for a given duration. This metric was meant to provide a straight-forward method of assessing workload, airspace complexity, and safety between different levels of automation.

Having established the overall number of separation violations in the previous section where aircraft passed within less than five nautical miles laterally and 1000 feet vertically, the results for the minimum separation metric build on those numbers by including the number of unique aircraft pairs that also passed within five to 10 nautical miles (nm) laterally and 1000 feet vertically. The total numbers for both of these measures are also subdivided by five minute increments according to the times at which they occurred. In this study, the final result for this metric affords a fuller picture of the way in which the conditions played out with respect to the ways in which the different modes of conflict resolution impacted the complexity and composition of the traffic flow across the three levels of traffic density. As this metric was intended to be a means of simple comparison, statistical analyses were not conducted. However, the trends and characteristics that came out of this comparison will be presented.

As seen in Figure 9, the three conflict resolution modes are presented for the 1X level of traffic. Because each of these three conditions involved the same two simulation scenarios, it is not surprising to see that the first five minutes are nearly identical across the three resolution modes in terms of the number of aircraft within 10 nm of one another. Each of the three conditions has at least one aircraft pair within 10 nm of each other within the first five minutes. However, by the 10 minute mark this number decreases, and differences between the resolution modes can already be seen with the fully automated mode showing no aircraft pairs within the 10 nm criterion while the manual and interactive modes did produce a few discernable number of pairs. Following this initial decrease, each condition exhibits a progressive increase in the number of aircraft pairs within 10 nm of one another. Of note here is that the manual condition appears to have a shallower rate of increase over time relative to the other two modes. The peak number of 3.33 is also less than the interactive and fully automated conditions' peaks of 4.44 and 5.00 aircraft pairs respectively. These numbers suggest that having to manage the resolution of conflicts manually somehow enabled the participants to achieve a more conservative and safe level of separation between aircraft.

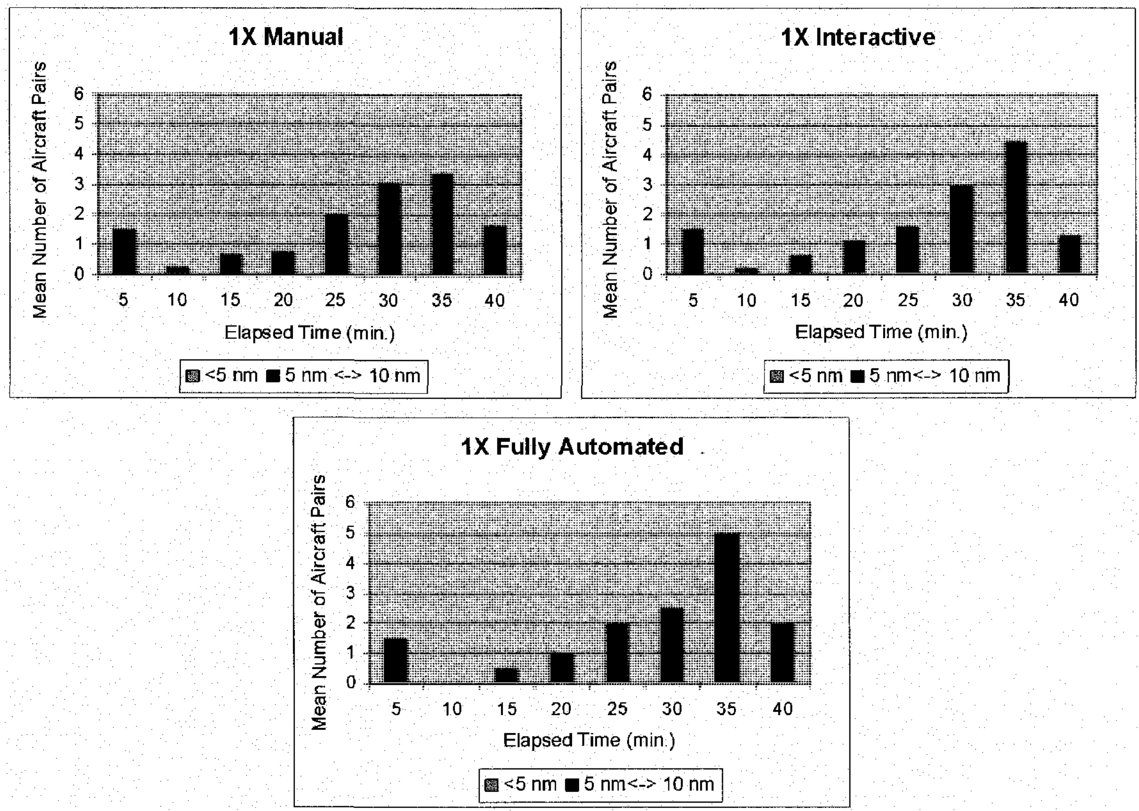


Figure 9. Minimum separation distances for the 1X level of traffic.

Progression from the 1X to the 2X level of traffic saw an increase in the overall number of aircraft pairs passing within 10 nm of one another (see Figure 10) due in part to the decrease in airspace available to accommodate the increase in traffic. In contrast to the results from the 1X level, differences between the resolution modes at the 2X level can be seen almost immediately. As opposed to the manual and fully automated conditions, the timeline for the interactive mode shows that separation violations were occurring within the first five minutes of the scenario. Despite this tenuous start, it turns out that the manual mode of conflict resolution resulted in the most frequently occurring number of separation violations as seen in Figure 10 where it is shown that violations occurred over four consecutive time steps starting from the 10 minute mark. Although this is a grave safety concern, it is interesting to note that despite those separation violations in the manual condition, much like at the 1X level, the overall count of aircraft pairs passing within 10 nm of one another remained low compared to the other two modes. And, once again, the manual condition had the smallest average peak at 13.72 compared to the 14.56 and 15.50 peaks of the interactive and fully automated conditions respectively. However, given the frequency of separation violations in the manual condition, it appears as though the interactive condition was the safest at the 2X level of traffic, all things considered. Despite having a slightly higher peak than the manual condition, the interactive condition actually had a slightly lower overall average number of aircraft pairs that exceeded the separation criterion. The average number for an entire run in the interactive condition was 61.25 whereas the average number in the manual condition was 62.31 and the fully automated condition was 70.00.

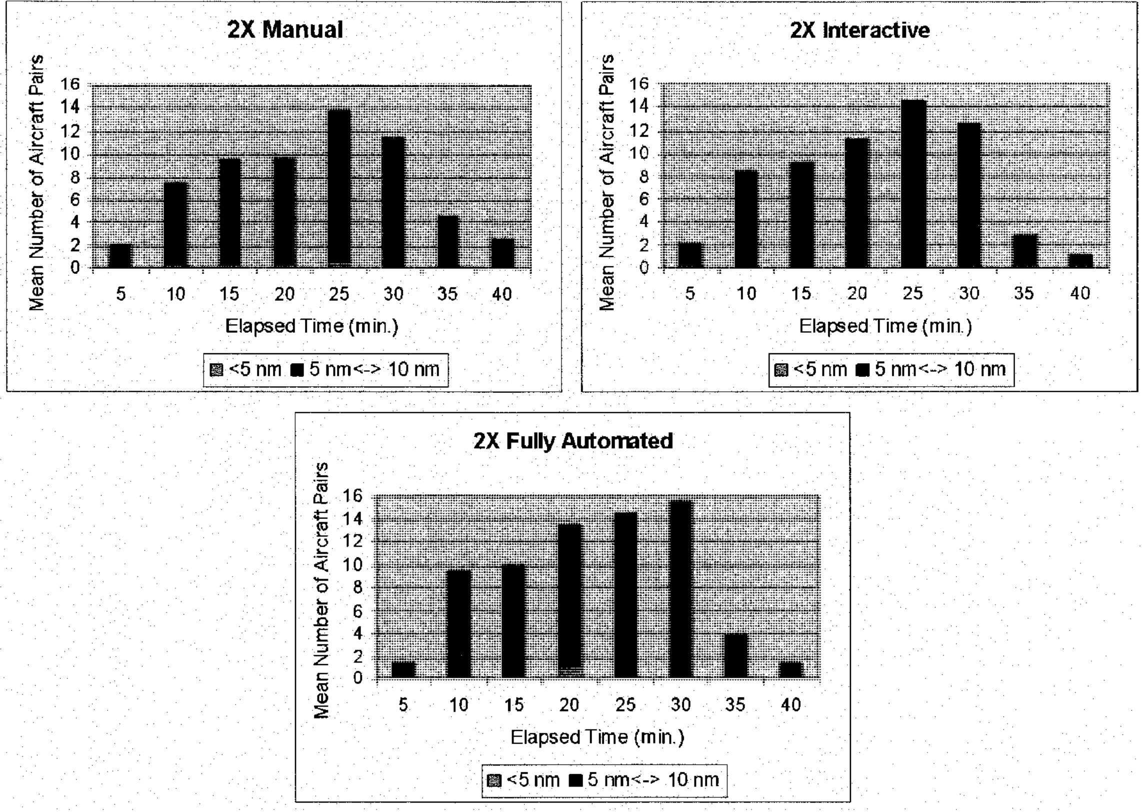


Figure 10. Minimum separation distances for the 2X level of traffic.

While it seemed as though the automated conflict resolution tool at the 1X level of traffic did not provide great benefits in terms of safety and complexity reduction, it did appear to do so at the 2X level of traffic. Interestingly, it seems as though the greatest benefit was seen when a human was in charge of using the tool selectively as opposed to the tool being applied in a completely autonomous manner.

As the 2X level of traffic showed an overall increase in the number of separation violations and aircraft pairs that passed within 10 nm miles and 1000 ft of one another, this trend saw a stark increase at the 3X level of traffic, particularly in the manual mode of conflict resolution. Although this trend was not surprising as the sheer volume of traffic essentially choked the airspace thus reducing the degrees of freedom for safe and conservative resolution maneuvers, it is interesting to note the stark differences between the different resolution modes at this level.

As shown in Figure 11, participants in the manual mode experienced a number of separation violations at each time step, with numbers at one point reaching 10 separation violations during a single five minute period. These kinds of numbers also meant that the overall number of aircraft pairs within 10nm was greatly increased. Figure 11 shows the timeline for the manual mode where the number of aircraft pairs increased dramatically to a point at which the peak number of 40.76 was achieved. This stands in contrast to the interactive and fully automated modes where the numbers of aircraft pairs barely exceeded 25. With respect to the latter conditions, much like at the 2X traffic level, the interactive condition showed the least average overall number of aircraft pairs within the separation criteria with an average of 121.81 pairs per run as opposed to the 125.00 pairs

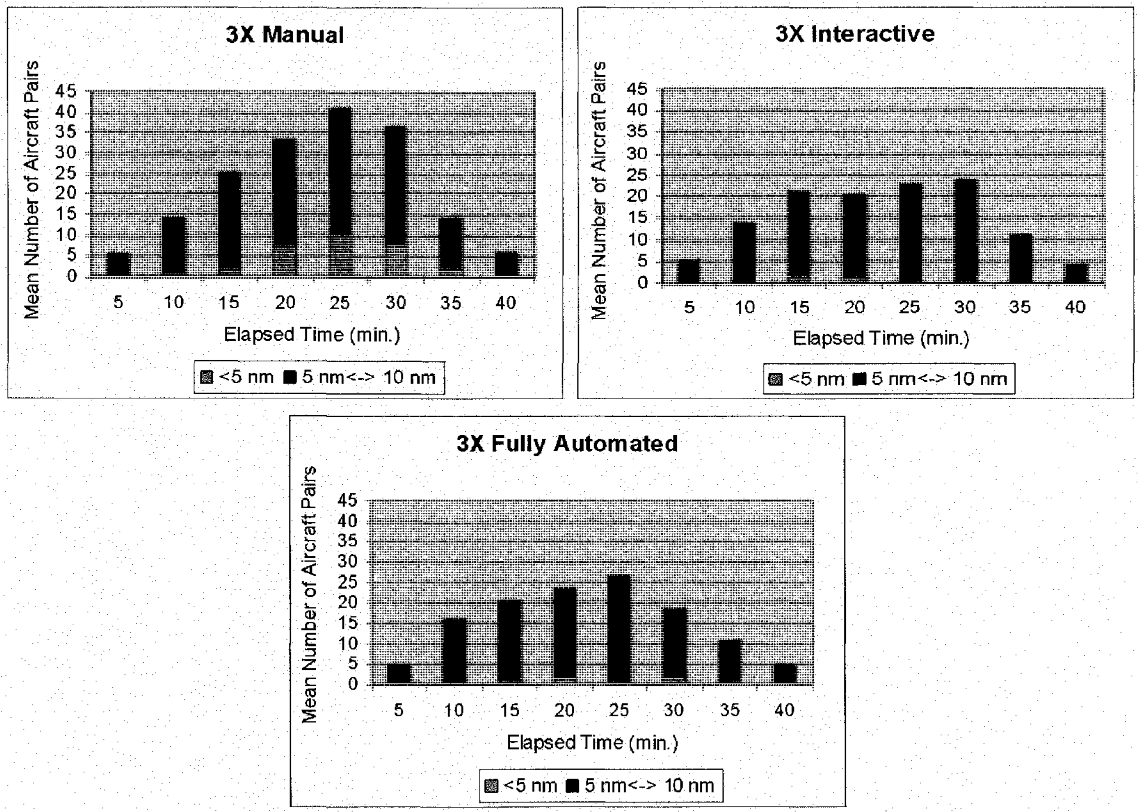


Figure 11. Minimum separation distances at the 3X level of traffic.

seen in the fully automated condition. However, as presented in the previous subsection, the fully automated condition had the fewest separation violations at this traffic level out of any of the conflict resolution modes. One point worth mentioning here is that particularly at the 3X level of traffic, it was interesting to see how greatly the results for the manual and interactive conditions differed. In this case, the interactive condition was much more similar to what was seen in the fully automated condition. Perhaps the results from the participants' usage characteristics of the automation presented in the Acceptability section will shed some light on this matter.

Efficiency

For this study, efficiency was defined in terms of four separate yet related measures. The first measure was the total delay added to the airspace system as a result of the conflict resolution clearances executed. The average delay per resolution was also included as part of the definition of efficiency. The next measure was the average lateral distance deviation, relative to the original flight path, per lateral resolution clearance. The final measure was the average vertical distance traversed by aircraft given altitude clearances.

Total Delay

Total delay was measured in seconds and was the result of both lateral and altitude conflict resolutions. The descriptive statistics for this measure show that for the manual condition's 1X level, instead of imposing delays, conflict resolutions actually afforded an average time savings of -28.25 seconds (*SD*= 185.94) (see Figure 12).

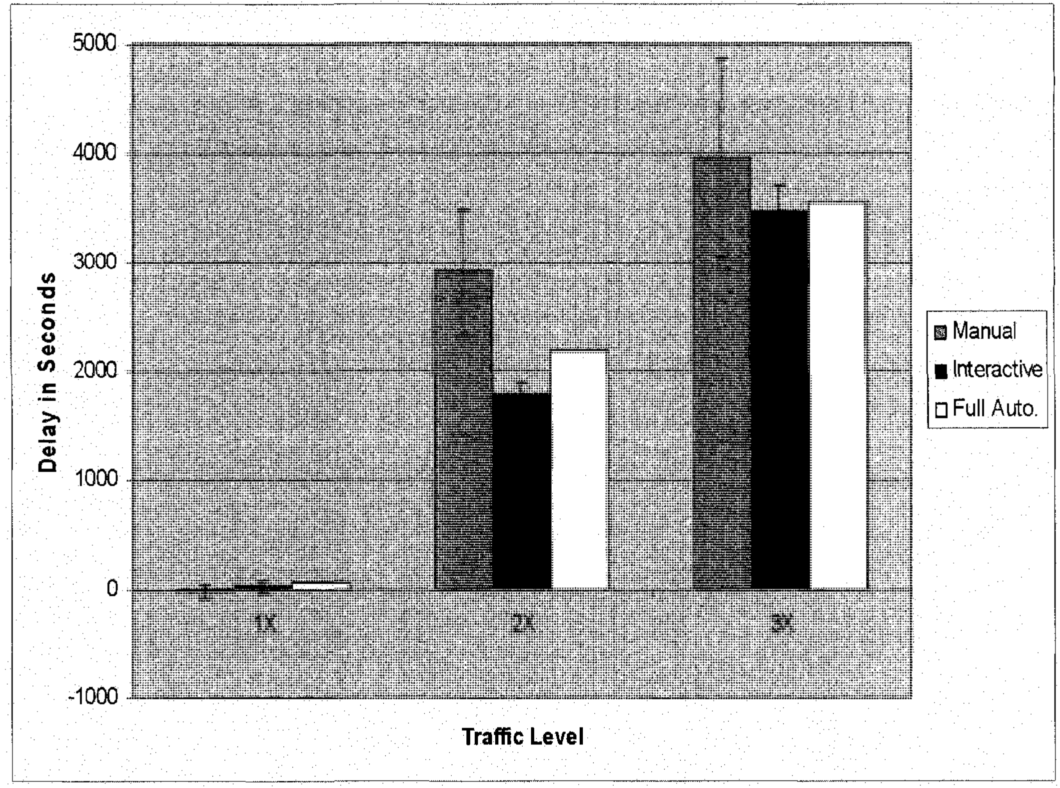


Figure 12. Total delay imposed by conflict resolutions in each condition.

This affordance of time savings related to conflict resolutions did not persist, however, as the traffic level was increased to 2X (*M*= 2908.06, *SD*= 1636.97). This trend continued to the 3X level of traffic where the conflict resolutions enacted resulted in greater average delays (*M*= 3954.81, SD= 2556.08).

Unlike the 1X traffic level in the manual conflict resolution mode, the interactive mode at 1X did actually result in delay with an average of 24.63 seconds (SD= 180.78).However, at the 2X level of traffic, the interactive mode imposed less delay than what was observed in the manual mode (M= 1770.38, SD= 346.96). The 3X level of traffic also saw the interactive mode providing less delay than the manual mode (M= 3476.13, SD= 668.31). Note that the standard deviation values for the interactive condition were also much less than what was observed in the manual condition.

The fully automated mode of conflict resolution turned out to have delay results that were between what was seen in the manual and interactive conditions with one exception; At the 1X level of traffic, the fully automated condition actually had the highest average amount of delay out of any of the resolution modes at 58.00 seconds (*SD*= 0.00). However, the 2X (*M*= 2180.50, *SD*= 0.00) and 3X (*M*= 3567.00, *SD*= 0.00) levels of traffic saw an average amount of delay that was less than what was observed in the manual conditions.

In terms of the delay aspect of efficiency, based on the descriptive statistics just presented, it appears as though at current day levels of traffic, the participants were able to provide the most efficient resolutions without the aid of the automated resolution tool. However, this was only the case at the 1X level. As the traffic levels progressed to 2X and 3X, it appeared as though the automated tool provided efficiency benefits relative to when the tool was not available in the manual condition. For the least amount of delay at these levels, however, the results showed that keeping the human involved in the process did have an advantage as the interactive conditions had less delay compared to when the application of the tool was fully automated.

Although the descriptive statistics showed that the interactive condition had the least delay at the higher levels of traffic while the manual condition had the least delay at current day levels, a further exploration of these differences was needed in order to understand their significance. To that end, a 2 (manual vs. interactive resolution mode) x 3 (1X, 2X, 3X traffic levels) repeated measures analysis of variance (ANOVA) was conducted on the total delay added to the system as a result of the conflict resolutions.

Despite the differences in total delay between the two resolution modes that were evident in their descriptive statistics, no significant main effect was found, $F(1, 7)= 1.76$, $p> .05$. However, there was a significant main effect for traffic level, $F(2, 14)= 45.68$, $p<.01$. This did not result in a significant interaction between the two variables though, $F(2, 14)= 1.70, p>.05$. The fact that there was a significant main effect for traffic level did not come as a surprise given the extreme differences between the three levels both in terms of the scenarios as well as the resulting numbers outlined in the descriptive statistics.

The preceding statistics referred to the measure of total delay imposed on the system as a result of all conflict resolutions given over the course of an experimental condition's run. This was an important measure to include as the total delay is, in the end, a quantifiable and translatable metric that is of definite interest to all stakeholders of the air transportation system. However, the results just presented do not provide a full picture of delay as each condition afforded the participants the ability to send up a different number of clearances based on time available and the complexity of the situation. As a result, some of the differences in delay observed thus far were simply the result of there being a fewer or greater number of clearances issued. Therefore, an additional measure of the average delay per clearance was included in order to equalize the various resultant delays across all of the study's conditions.

Average Delay

The descriptive statistics for the average delay characterized in Figure 13 show that in the manual condition's 1X level of traffic, the average delay per clearance uplink was still minimal (*M*= 0.06, *SD*= 34.76). At the 2X level, the average delay per resolution increased dramatically to 70.49 seconds (*SD*= 40.42). The increase continued to the 3X level, but at a far lesser rate (*M*= 80.62, *SD*= 63.88).

For the interactive conditions, the average delay per clearance at the 1X level was greater than its manual condition counterpart (*M*= 6.65, *SD*= 32.56). The 2X level showed an increase to an average of 35.25 seconds (*SD*= 6.42). Interestingly, the 3X level only saw a slight increase in average delay over the 2X level (*M*= 37.03, *SD*= 5.40). At the 1X level of traffic for the fully automated condition, the average delay turned out to be considerably more than what was observed in the other two resolution modes with an average of 15.95 seconds (*SD*= 0.00) per uplink. At the 2X and 3X levels, however, the delay looks somewhat more in line with what was seen for the interactive conditions

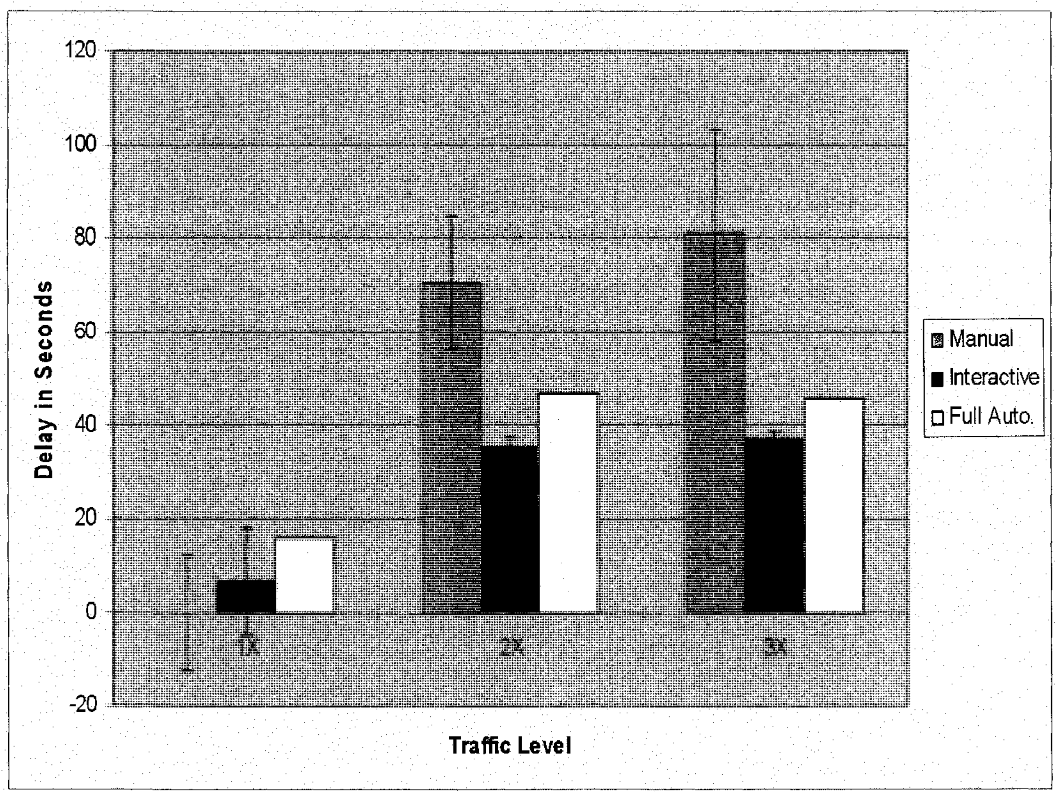


Figure 13. Average delay per resolution in each condition.

with averages of 46.83 seconds (*SD*= 0.00) and 45.43 seconds (*SD*= 0.00) respectively.

The descriptive statistics just presented seem to show that at the 1X level of traffic, the manual mode of conflict resolution appeared to have the least impact on delay relative to the interactive and fully automated modes. However, this benefit was short-lived as the manual mode showed the highest average delay out of any of the modes at the 2X and 3X levels. Conversely, the interactive mode appeared to have the least amount of delay associated with each clearance at these levels of traffic.

To further probe the differences in average delay between the manual and interactive conditions, a two-way, repeated measures ANOVA was conducted with resolution mode and traffic levels serving as the variables. In terms of resolution mode, despite the differences highlighted in the descriptive statistics, there did not appear to be a significant main effect, $F(1, 7)= 3.41, p> .05$. There was, however, a significant main effect for traffic level, $F(2, 14)= 30.82, p< .01$. Unlike the results observed for total delay, a significant interaction was found between resolution mode and traffic level for average delay, $F(2, 14)= 5.60, p< .05$. Despite the significant interaction that was found, subsequent paired t-tests of the manual and interactive resolution modes at each level of traffic failed to find any significant differences. While initially this might come as a surprise, a look back at the standard deviations presented in the descriptive statistics shows that the variance in many of the conditions was quite high, which ultimately served to weaken the effects of the observed mean differences between resolution modes.

Lateral Distance

Another measure of efficiency analyzed in this study was the added lateral distances flown, relative to the original flight path, by aircraft given lateral conflict resolutions. In this case, the average lateral distance flown per clearance was the measure of comparison because a cumulative measure would not translate to anything meaningful. Finally, distance was measured in nautical miles (nm) and the comparisons involved resolution mode and traffic level much like in the analysis of delay.

The descriptive statistics for the manual condition show that at the 1X level of traffic, participants were able to reduce the amount of distance flown as the average added distance was actually negative (*M*= -1.25, *SD*= 4.97) (see Figure 14). This finding is similar to what was observed for the total added delay outlined in the previous section where the 1X manual condition resulted in time savings. There is, however, some concern of a possible confound influencing the results for this condition in terms of participant behavior that will be discussed in the Discussion section. As the traffic level increased to the 2X level in the manual condition, there was an expected increase in the added lateral distance in that the average total distance added was 4.38 nm (*SD*= 3.20). The distance increased once again as the traffic level reached 3X (*M*= 11.58, *SD*= 5.56). For the interactive mode of conflict resolution, the 1X level of traffic showed greater average distances flown by each conflict aircraft than in the manual condition with an average lateral distance of 1.25 nm (*SD*= 3.65) per resolution clearance.

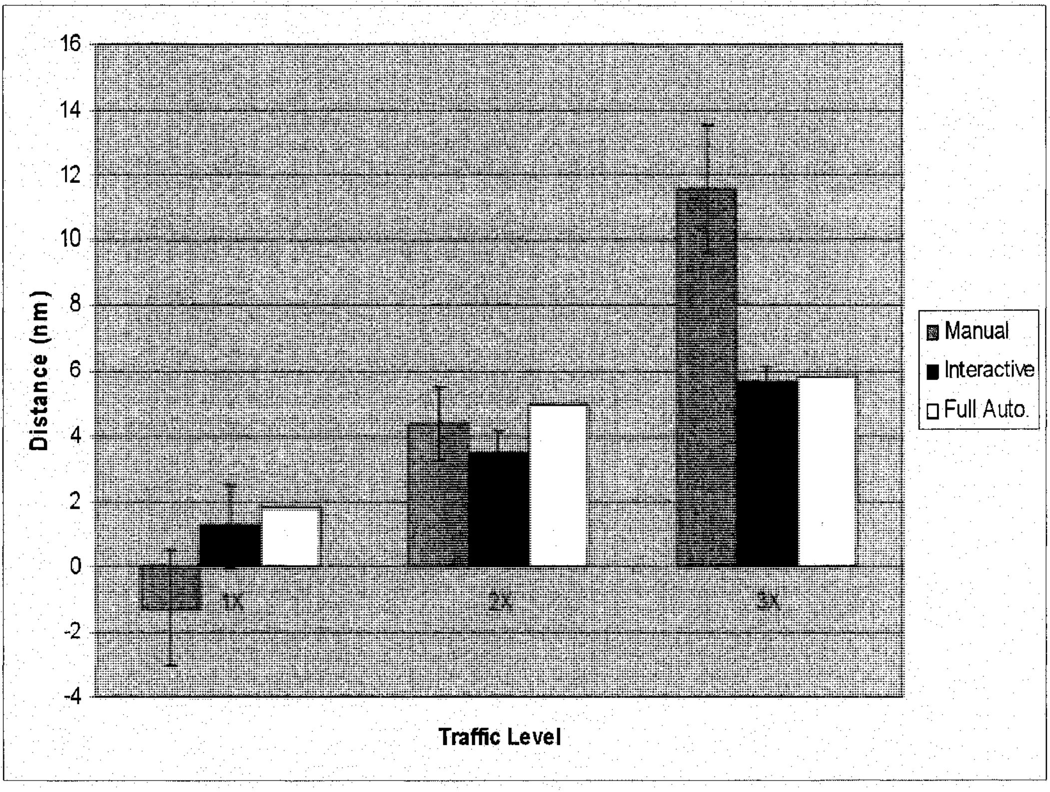


Figure 14. Average lateral distance per clearance in each condition.

However, at the 2X level, the interactive condition had a lower average lateral distance of 3.49 nm (*SD*= 1.97). At the 3X level of traffic, the difference between the two conditions was even more pronounced with an average of 5.66 nm (*SD*= 1.22) per clearance.

At the 1X and 2X levels of traffic, the fully automated mode of conflict resolution resulted in the greatest average lateral distance per clearance out of all resolution modes with values of 1.84 nm (SD= 0.00) and 4.97 nm (SD= 0.00) per resolution respectively. However, the 3X level of traffic showed that the fully automated condition only had a slightly larger average lateral distance than the interactive condition, which had the least amount of lateral distance added per clearance (M= 5.84, SD= 0.00).

These descriptive statistics suggest that giving the participants access to the automated conflict resolution tool allowed them to construct lateral conflict resolution clearances that were more efficient at the higher levels of traffic than when they did not have the tool. Likewise, it appeared as though the participants having the tool and being able to judge and apply it selectively resulted in more efficient lateral resolutions than when the automated tool was allowed to operate independently.

Given the differences observed between the manual and interactive modes of resolution at each level of traffic, it was important to further investigate these differences in order to understand their significance. To that end, a two-way, repeated measures ANOVA was conducted on the average lateral distance per clearance. With regard to resolution mode, a significant main effect was not found, $F(1, 7)= 3.43, p> .05$. A significant main effect was found, however, for traffic level, $F(2, 14)= 33.91, p <01$. A significant interaction was found between the resolution mode and traffic level, $F(2, 14)=$

8.03, p <01. As highlighted in the descriptive statistics, the most profound differences between the resolution modes in average lateral distances were found at the 3X level of traffic. Subsequent paired t-tests confirmed that the differences between the manual and interactive modes were only significant at the 3X traffic level, $t(7)= 3.56$, $p<.01$. A look back at the descriptive statistics showed that the interactive mode of conflict resolution provided the most efficient lateral resolutions at the 3X level of traffic with an average distance of 5.65 nm compared with the 11.58 nm seen in the manual condition.

Vertical Distance

The final measure of efficiency was the average vertical distance traversed by aircraft in each vertical conflict resolution that was implemented. The distances used in the analysis were defined in units of feet (ft) and were taken in absolute terms without regard to the directions –up or down- that the resolutions involved the aircraft flying.

The descriptive statistics for the manual resolution mode show that at the 1X level of traffic, the average vertical distance for each vertical clearance was 3229.78 ft (*SD*= 1374.26). As shown in Figure 15, the average distance decreased slightly as the traffic level increased to 2X with an average distance of 3151.25 ft (*SD*= 1014.39) per vertical resolution. This decrease continued to the 3X level where the average distance was 2761.55 ft (*SD*= 826.86).

The interactive conditions followed similar trends as seen in the manual condition but with less distance associated with each resolution. For example, at the 1X level of traffic, the average vertical distance flown was found to be 2022.00 ft (*SD*= 1298.03).

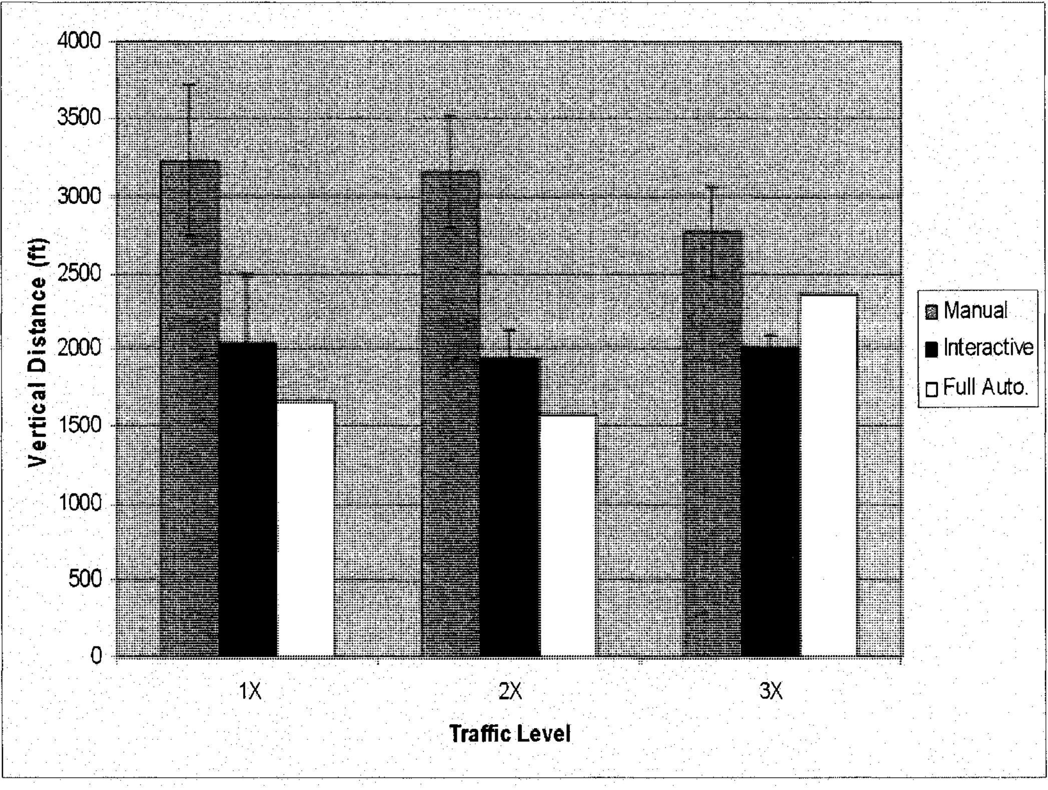


Figure 15. Vertical distance per vertical resolution in each condition.

The 2X level resulted in an average of 1944.81 ft (*SD*= 507.80) with a further decrease at the 3X level of traffic (*M*= 1996.82, *SD*= 242.55).

The fully automated condition had the least vertical distances per clearance at the 1X and 2X levels of traffic with respective averages of 1666.67 ft (*SD*= 0.00) and 1572.71 (*SD*= 0.00). At the 3X level, however, the average distance was 2363.39 ft (*SD*= 0.00), which was actually higher than what was observed in the comparative interactive condition.

Based on these descriptive statistics, it seems as though the fully automated mode of conflict resolution allowed the tool to find vertical resolutions that required the least amount of vertical change at the 1X and 2X levels. However, at the 3X level, the interactive condition added the least amount of vertical distance per clearance, which suggests that the participants were somehow better able to use the automated tool at their discretion to find more efficient vertical resolutions when the traffic reached such an elevated level.

The differences observed in the descriptive statistics required further analysis in order to see the significance of each condition's effect on average vertical distance. Once again, a two-way, repeated measures ANOVA was conducted on average vertical distance with resolution mode and traffic level as the variables. With respect to resolution mode, it turns out that there was a significant main effect, $F(1, 7)= 8.21, p<.05$. However, traffic level did not have a significant main effect, $F(2, 14)= .45, p> .05$. There was no significant interaction between the two variables, $F(2, 14)= .28, p>.05$. Given the differences between the manual and interactive conditions observed in the descriptive statistics, a significant main effect was not a surprising result with the interactive condition clearly enabling more efficient vertical clearances.

Workload

Following efficiency, the subsequent focus of analysis was workload. Measures of workload were taken via two different sources. The first source of workload data was derived from responses to prompts of an emulated workload assessment keypad (WAK) in MACS taken during each run at five minute intervals. From these responses, the average and peak workload data was used for analysis. The second source of workload data was taken from post-run questionnaires that asked for the peak workload experienced during the previous run.

For the workload ratings given during the run, participants rated their subjective workload over the preceding five minutes on a scale from one to seven, with seven representing the highest workload possible. The first point of analysis for this data was the average workload experienced for each condition. As in the previous results detailed thus far, the sample size was *N*= 8 due to the fact that one participant did not complete both runs of the 3X manual condition. Additionally, the workload analysis will obviously only include the manual and interactive conditions at each of the three traffic levels as the participants were not involved in the fully automated condition.

As shown in Figure 16, the obvious trend was an increase in average workload as the traffic level was increased. However, the rate of increase was not the same between the manual and interactive conditions. To be more specific, the descriptive statistics for the average workload in the manual condition show that at the 1X level of traffic, the

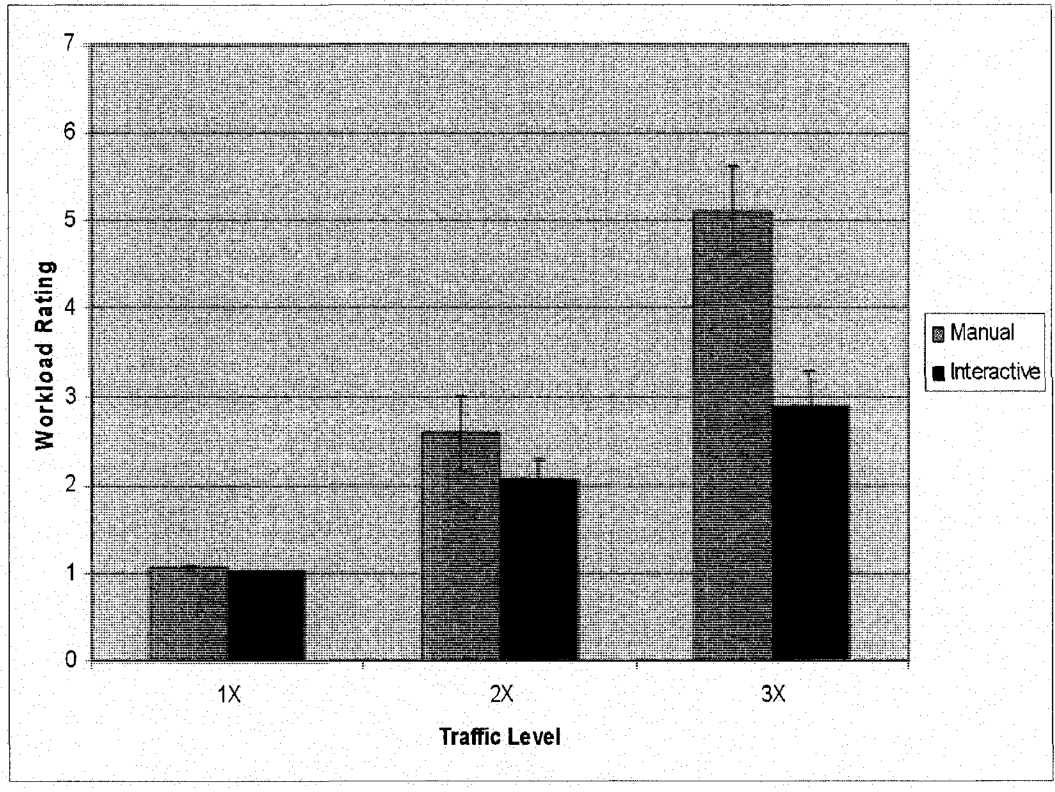


Figure 16. Average workload ratings reported during each run.

average workload was nearly as low as possible with an average of 1.05 (*SD*= 0.10). Average workload appeared to increase at the 2X level of traffic as participants reported an average of 2.57 (*SD*= 1.21). Not surprisingly, the largest increase in workload came at the 3X level of traffic where an average of 5.12 (*SD*= 1.44) was recorded.

The interactive condition showed slightly lower average workload levels at the1X and 2X levels of traffic with mean reports of 1.01 (*SD*= 0.03) and 2.06 (*SD*= 0.64) respectively. At the 3X level of traffic, however, the average recorded workload was 2.88 (*SD*= 1.16). This represented a rather large difference between the manual and interactive modes of conflict resolution, which suggests that the aid of the automated conflict resolution tool begins to provide the greatest benefit in terms of workload reduction in particularly highly dense and complex traffic situations as seen at the 3X level of traffic in this study.

To investigate whether or not there was any significance to the differences observed in the descriptive statistics, a two-way, repeated measures ANOVA was conducted on average workload in which the resolution modes of manual and interactive were compared with the three levels of traffic.

Results of the ANOVA revealed significant main effects for resolution mode, $F(1, 7)= 31.21, p< .01$, and traffic level, $F(2, 14)= 48.60, p< .01$. A significant interaction between the two was also found, $F(2, 14)= 7.18, p< .01$. After conducting subsequent *t*-tests to isolate the differences in workload, it was found that, as seen with the descriptive statistics, the point at which there was a significant difference between the two resolution modes was at the 3X level, *t*(7)= 3.79, *p*< .01. Again, the interactive mode showed the greatest benefit in reducing workload relative to the manual mode.

The results for average workload were included in order to get a general sense of workload experienced by the participants throughout each condition's runs. However, as always with averaging, some of the more interesting results get diluted and their importance is lost. In this case, it was felt that the peak workload ratings across the conditions would be important to analyze as they would give a truer sense of the impact that use of the automated tool had on workload. To that end, the highest reported workload rating per run was used in the following analysis. One item to note before proceeding is that as the previous analyses used a sample size of *N*= 8, the analysis of peak workload included the addition of the ninth participant due to the assumption that at the point that the individual decided to withdraw from the 3X manual runs, the workload rating was basically equivalent to the highest possible rating of seven and was included in the analysis as such.

The descriptive statistics for the manual condition at the 1X level show that the peak level of workload hardly ever exceeded the lowest possible rating of one (*M*= 1.39, *SD*= 0.60) (see Figure 17). The 2X level showed a noticeable increase with an average peak rating of 3.67 (*SD*= 1.48). The 3X level of traffic resulted in by far the highest average peak ratings at 6.39 (*SD*= 1.32). The interactive conditions followed the same trends observed in the manual Condition, but with lower peak workload ratings. The 1X level of traffic was similar to what was seen in the manual condition with an average peak rating of 1.06 (*SD*= 0.17). The 2X level showed an increase in peak workload that

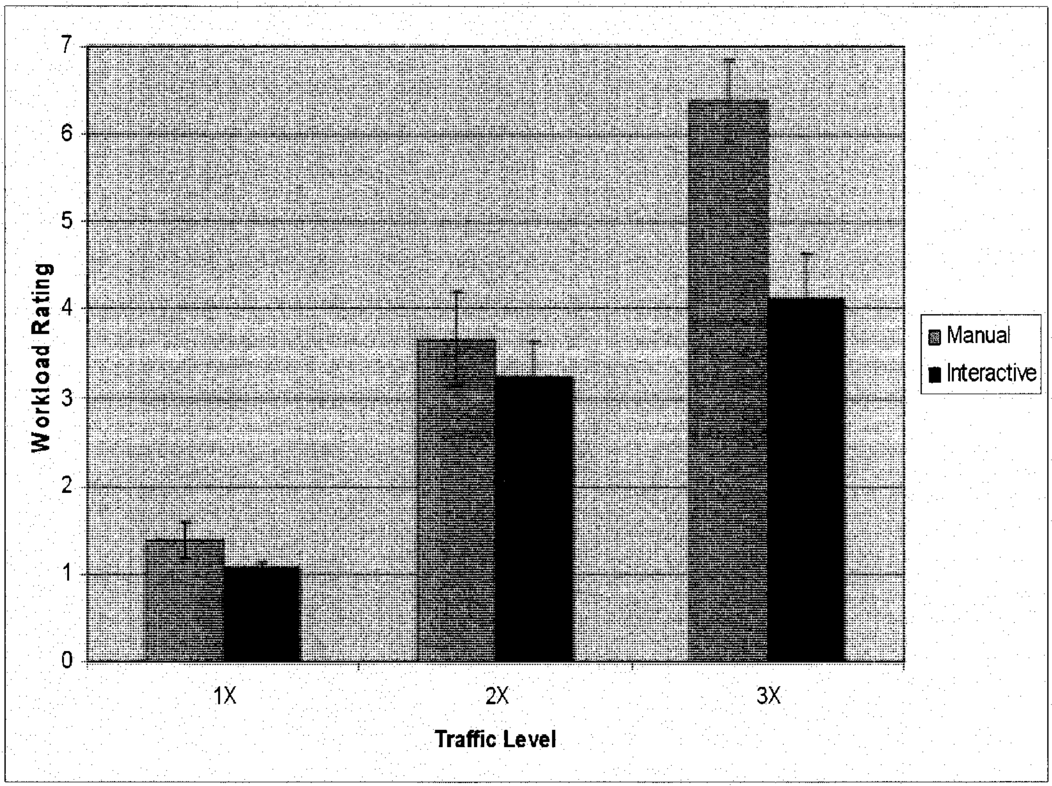


Figure 17. Peak workload ratings taken during each manual and interactive run.

was still slightly less than its manual condition counterpart (*M*= 3.22, *SD*= 1.20). The greatest difference between the resolution modes was seen, as in other measures, at the 3X level of traffic with the interactive condition's average peak ratings being 4.11 (*SD*= 1.50).

As before in the average workload analysis, a two-way, repeated measures ANOVA was conducted on peak workload in order to test the significance of the differences between the conflict resolution modes across each of the three levels of traffic. The only difference here was that as the analysis involved peak workload ratings, all nine participants were able to be included. Results of the ANOVA showed a significant main effect for resolution mode, $F(1, 8)= 26.65$, $p< .01$, as well as traffic level, $F(2, 16)= 95.26$, $p< .01$. A significant interaction between the two variables was also found, $F(2, 16)= 4.79$, $p< .05$. To further explore the nature of the significant interaction, multiple paired *t*-tests were conducted using the resolution modes for comparison at each level of traffic. Much like in the previous analysis of the average workload, the results of the *t*-tests showed that it was at the 3X level of traffic that the benefit of the automated tool in reducing workload was maximally realized with a significant result of $t(8)= 5.35$, $p< .01$.

As mentioned previously, a secondary source of workload data was gained from post-run questionnaires that asked participants to rate their peak workload, on a scale from one to seven, over the entire previous run. Although this information could be gained directly from the runs as was just presented, it was thought that asking for peak workload responses following each run would allow for more informed, holistic ratings of peak workload after the participants had time to reflect on the run that they had just completed.

Figure 18 shows that the results for peak workload in this case were very much similar to what was observed in the previous analysis of peak workload. The descriptive statistics for the manual conditions showed that at the 1X level, the average peak workload was 1.28 (*SD*= 0.44). Average peak workload at the 2X level was 4.39 (*SD*= 1.50). The 3X level of traffic resulted in extremely high peak ratings with an average of 6.94 (*SD*= 0.17). This basically meant that at some point during the run, each participant had reached workload levels that were nearly intolerable.

The interactive conditions showed slightly lower peak ratings than the manual condition particularly at the 1X and 2X levels of traffic with results of 1.00 (*SD*= 0.00) and 3.17 (*SD*= 1.00) respectively. The 3X level of traffic produced the largest difference between the resolution modes with the average peak workload rating being 4.89 (*SD*= 1.65). Overall, perhaps the most interesting item to note from these descriptive statistics was the extreme magnitude of the 3X manual condition's peak ratings of nearly the maximum of seven for all nine participants.

For analysis of the inferential statistics, the same two-way, repeated measures ANOVA was conducted with essentially the same results: a significant main effect for resolution mode, $F(1, 8)= 59.80$, $p< .01$, a significant main effect for traffic level, $F(2, 16)= 142.75$, $p< .01$, and a significant interaction between the two variables, $F(2, 16)= 4.70$, $p< .05$. Following the discovery of the significant interaction, paired *t*-tests were conducted, and it was here that a noticeable difference emerged between the previous

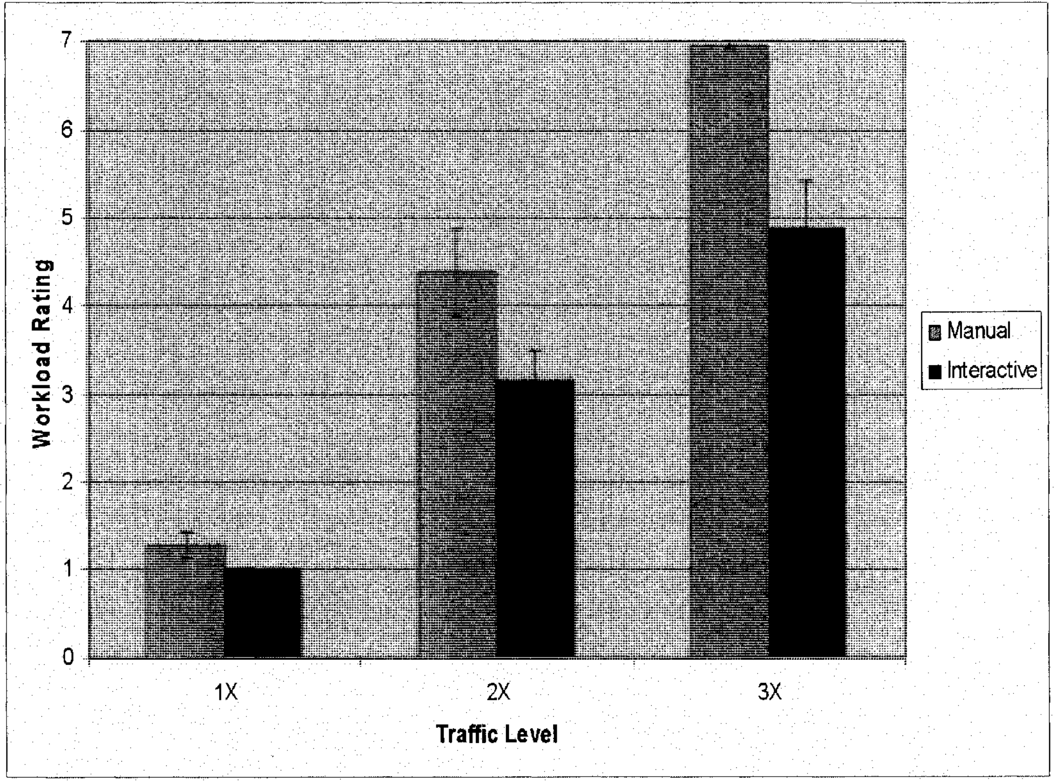


Figure 18. Peak workload ratings from post-run questionnaires.

peak workload analysis and the current analysis. In the previous analysis, the point at which a significant difference was found between the manual and interactive modes was at the 3X level of traffic. In the current analysis of peak workload ratings gathered from questionnaires, results from the paired *t*-tests showed that there was a significant difference between the two modes at the 2X level of traffic, $t(8)= 4.05$, $p<.01$, as well as at the 3X level of traffic, $t(8)= 3.79$, $p< .01$. Another look at the descriptive statistics shows that it was the interactive mode of conflict resolution that served to reduce workload levels. This provides further support to the earlier results in that use of the automated conflict resolution tool by the participants provided an advantage in terms of workload over conditions where such aid was unavailable.

Acceptability

The final measure addressed in the current study was the acceptability of the automated conflict resolution tool's suggested resolutions. Acceptability was determined through two different means, one being objective and the other subjective. The objective measure of acceptability was obtained from the participants' usage characteristics of the automated resolution tool. The subjective measure of acceptability was derived from responses to an item on the post-run questionnaires given after completion of each run in the interactive conditions.

The data related to the objective measure of acceptability was gained by looking at the relationships between the number of times the automated conflict resolution tool was used relative to the overall number of resolutions that were attempted, and then contrasting the number of times the tool was used with the number of times that the suggested resolution was modified or rejected. The latter metric served as the proxy for acceptability as it was thought that the use of a resolution suggested by the algorithm meant that it was, for whatever reason, deemed acceptable by the participant. Conversely, rejection of the suggested resolution through either its modification or cancellation was considered an indication of unacceptability.

The following analysis was intended to be used more for the general understanding of the automated tool's acceptability and the trends associated with how it may have been impacted differently across the 1X, 2X, and 3X levels of traffic. It was not intended to discover the statistical significance. Consequently, the results presented from this analysis will be purely descriptive in nature. Additionally, as this portion of the analysis solely considers the use of the automated conflict resolution tool by the participants, all results presented are in relation to the interactive conditions only.

Figure 19 presents the mean number of resolutions attempted, the number of times the automated resolution tool was called upon, and the number of times the suggested resolution was modified. Given the differences between traffic levels, it is not surprising to see that the number of attempted resolutions increases steadily as the traffic level increases. To be more specific, the average number of attempted resolutions at the 1X level of traffic was 7.44 (*SD*= 1.33). Of those attempted resolutions, the average number of times that the automated conflict resolution tool was used was 5.44 (*SD*= 2.10). After using the tool and reviewing the suggested resolution, the average number of times that the resolution was chosen to be modified was a minimal 0.50 (*SD*= 0.66).

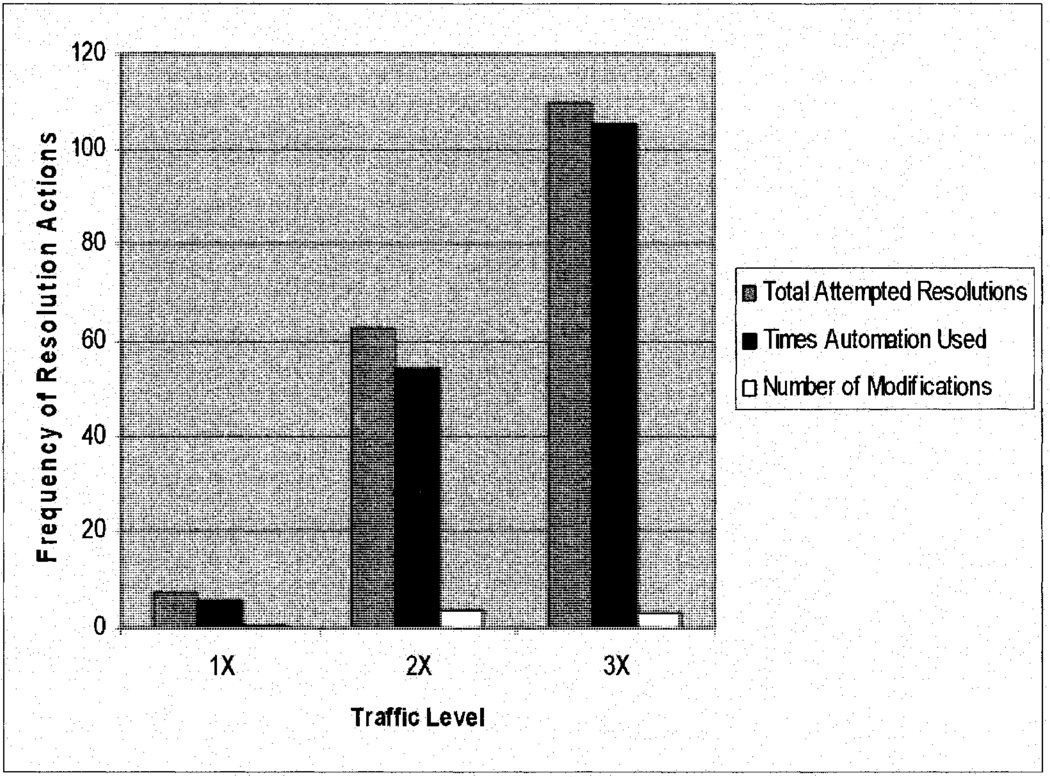


Figure 19. Automated conflict resolution tool usage characteristics.

Moving to the 2X level of traffic resulted in an average of 62.78 (*SD*= 8.62) attempted resolutions, a sharp increase over what was seen at the 1X level. Out of those attempted resolutions, the average number of times that the automated tool was called upon was 54.28 (*SD*= 7.89), from which there were an average of 3.56 (*SD*= 3.85) modifications to the suggested resolution. The 3X level of traffic, of course, showed the highest values for each of these measures with an average of 109.33 (*SD*= 12.33) attempted resolutions, an average of 105.22 (SD= 12.26) times that the automated tool was used, and an average 3.06 (*SD*= 2.54) modifications.

From the results just presented, it appears as though as the traffic level increased, so did the use of the automated conflict resolution tool. On the other hand, it also appears that with the increase in the tool's usage at the progressive levels of traffic, there was a decrease in the number of modifications to the resolutions by the automation.

To get a clearer picture of these trends and to neutralize the numerical differences brought about by the inherent differences between traffic levels, percentages of tool usage and resolution modifications were calculated. Figure 20 presents the percentages just referred to, and as one can see, the trends identified by the total values remained consistent. As the traffic level increased, so did the percentage of the time that the automated tool was used. More specifically, at the 1X level of traffic, the tool was used 73.13 % of the time. As the traffic increased to the 2X and 3X levels, the tool's usage increased to 86.46% and finally 96.24% of the time respectively. As the frequency at which the automated tool was used increased, the number of modifications decreased.

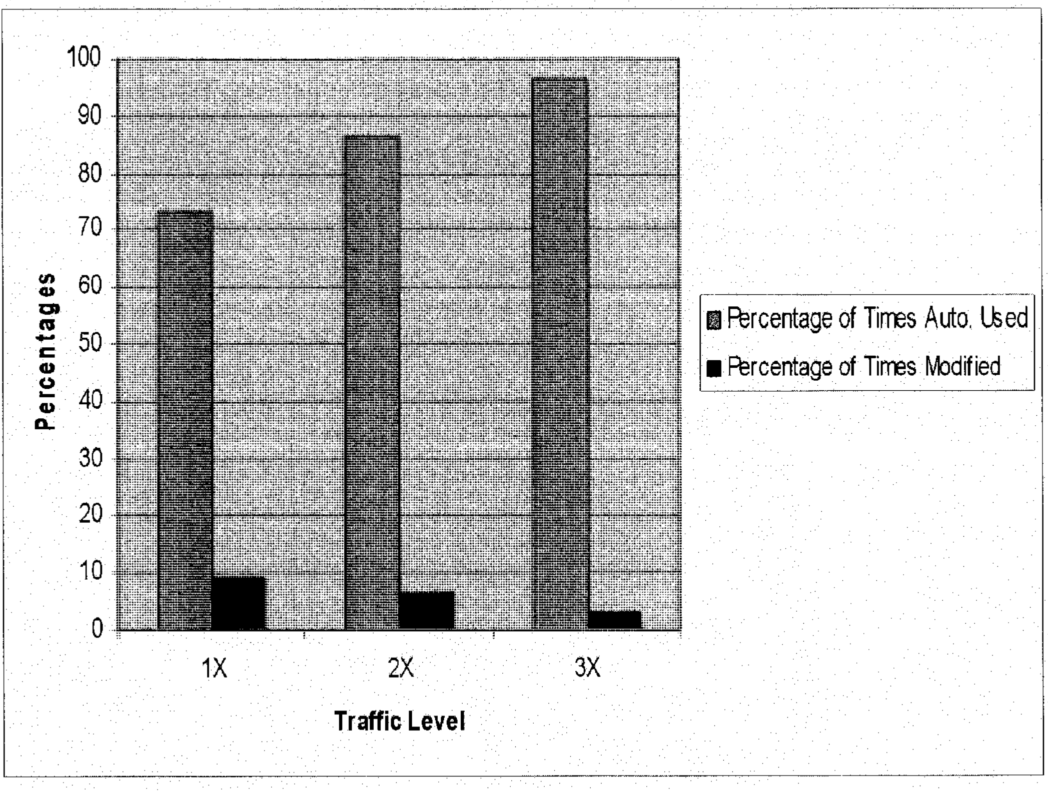


Figure 20. Percentages of times the automation was used and its resolutions modified.

At the 1X traffic level, the percentage of suggested resolutions that were modified was 9.18%. At the 2X level the percentage decreased to 6.55%, and finally at the 3X level the percentage of modifications fell to 2.90%.

Given the interpretation of acceptability previously outlined for these results, it appears as though the resolutions suggested by the automated tool were increasingly acceptable as the traffic density increased. The reasons for this are not entirely clear; however some possibilities will be discussed in the Discussion section.

The final measure of acceptability was subjective in nature and consisted of the participants' responses to an item on the post-run questionnaire following each interactive condition's run. The response was to a question that asked, "How acceptable do you feel the suggested conflict resolutions from the automation were?" This response was on a scale from one to seven with one representing the lowest level of acceptability and seven representing the highest level of acceptability.

Figure 21 presents the average responses to the acceptability question for each of the three levels of traffic. Right away one can see that there is an obvious decline in the reported acceptability of the automation's suggested resolutions as the traffic levels increased. Descriptive statistics across the traffic levels show that the acceptability ratings were highest at the 1X level with an average of 6.61 (*SD*= 0.99). The acceptability at the 2X level declined slightly to an average of 6.28 (*SD*= 1.00) followed by a further decline at the 3X level to 5.72 (*SD*= 1.41). These results show that for some reason, as the traffic levels increased and the airspace became more complex, the

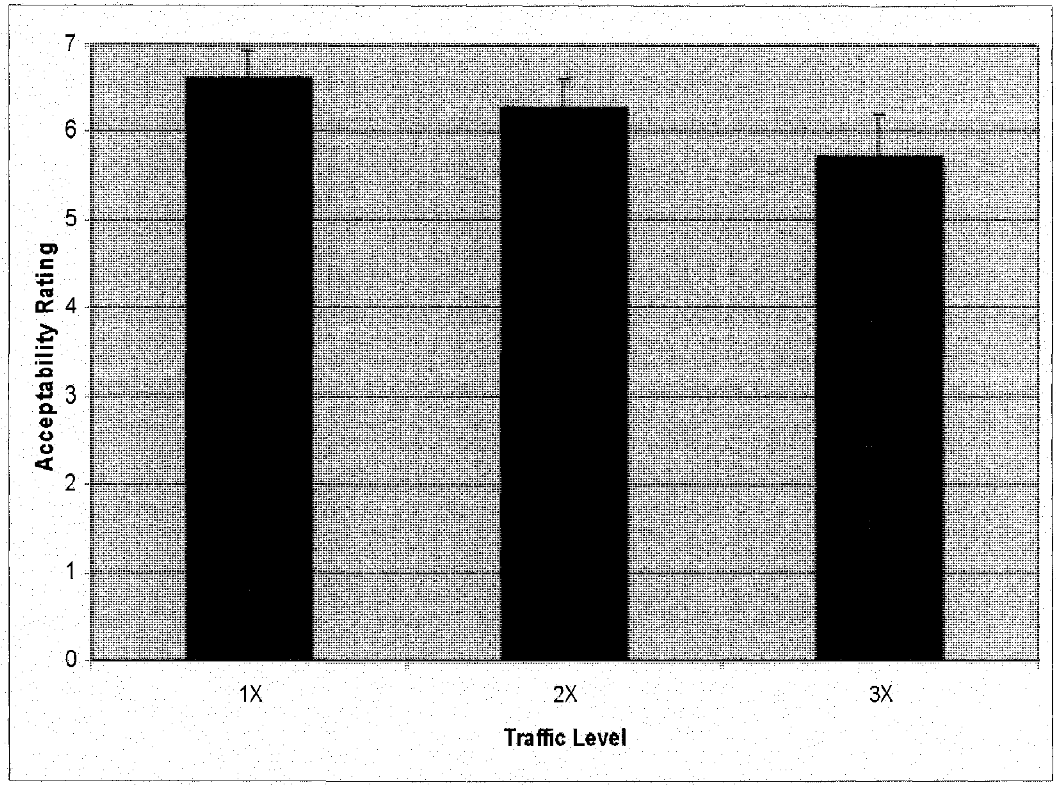


Figure 21. Average acceptability ratings from the post-run questionnaire.

behavior of the automated conflict resolution tool was such that the participants felt that the resolutions that it was suggesting became less acceptable to them.

To see whether or not the differences between the traffic levels were significant, a one-way, repeated measures ANOVA was conducted, with the three traffic levels serving as the variable levels. Results for this inferential statistic showed that the differences in acceptability between the traffic levels were significant, $F(2, 16)= 8.21, p< .01$. Paired *t*-tests with an adjusted Bonferroni procedure revealed that average acceptability ratings differed significantly between 1X and 3X levels, $t(8)= 3.25, p< .05$, as well as between the 2X and 3X levels, $t(8)= 3.16, p< .05$, but not between the 1X and 2X levels, $t(8)= 1.63, p> .05$.

According to the results just outlined, the resolutions suggested by the automated tool were least acceptable at the 3X level of traffic. However, a look back at the average acceptability rating shows that it was 5.72 out of a possible 7. This rating is still rather high despite being the lowest. In fact, the ratings across all levels of traffic suggest that the resolutions provided were acceptable overall. However, one might notice that the trend in this aspect of the analysis of decreasing acceptability is contrary to what was observed in the previous analysis of acceptability. In that instance, the interpretation was that acceptability increased as the traffic level increased. This is an interesting point to address and is deserving of further discussion in the following section.

DISCUSSION

The purpose of this study was to analyze the conflict resolution capabilities of humans and automation across three progressive levels of traffic. The focus of the analysis was on the impact that the different modes of conflict resolution had on the measures of system performance involving safety and efficiency, as well as the measures of workload and acceptability. The results just presented showed that the use of the automated conflict resolution tool had a beneficial impact on system performance and workload especially as the level of traffic increased. For example, the results for safety indicated that the use of the automation allowed for a significantly lower number of separation violations as early as the 2X level of traffic and continuing to the 3X level when compared to the results for the manual mode of conflict resolution. The automation also provided benefits for the measures of efficiency with significantly lower amounts of delay and lateral deviated distances associated with the conflict resolutions. Interestingly, while the differences in these efficiency measures between the manual and automated conflict resolution modes increased with the level of traffic, the opposite was observed for the measure of vertical distance. In addition to the system performance measures, the automation also had a beneficial impact on the workload reported by the participants. Once again, the benefits provided by the automation were most apparent particularly at the 3X level of traffic. Finally, the behavior and performance of the automated conflict resolution tool was generally acceptable to the participants though these results were not as easily interpretable. The results from this study have provided a number of topics for discussion as they relate both to the automation itself and the larger issues at hand for the national air transportation system. These topics will be discussed in turn with their implications incorporated into the discussion. This will be followed by the limitations of this study and the future directions that research could take in building upon what was accomplished through this initial study.

Safety

In terms of safety, although the number of separation violations was quite different among traffic levels and conflict resolution modes, the overall results were rather distressing. For instance, at the current day, 1X baseline traffic level there was one separation violation per manual and interactive resolution mode condition. While this might not sound like much, the gravity of a separation violation is such that even one occurrence is cause for alarm. In this case, with nine participants working a 1X scenario twice, the time that it took to observe a separation violation was only nine hours of operations. Keep in mind that this was confined to a specific airspace. If these numbers were to be translated to the national scale the frequency of separation violations would be much higher.

However, as detailed earlier, the current day system is much safer than these results would suggest. Therefore, the separation violation results for at least the 1X conditions should be interpreted cautiously. Perhaps one positive sign was that the fully automated condition for the 1X level of traffic did not produce any separation violations as was expected. The fact that there were any separation violations at this level in both the manual and interactive modes of resolution came as a bit of a surprise. It should be noted, however, that even though there was a separation violation in the interactive condition, participants were not forced to use the automated tool for each conflict that they encountered. In fact they did not use the tool as often at the 1X level in particular. This makes it a bit difficult to attribute the occurrence of the separation violation at this level to the tool per se. Nonetheless, it did occur in an interactive condition and was a surprise.

Less surprising, though, was that the number of separation violations was significantly higher at the 2X and 3X levels of traffic. Nowhere was this more apparent than at the 3X level of traffic where the mean number of separation violations in the manual condition was a staggering 28.19. This number was clearly unacceptable and provided further proof and support for drastic changes in the way air traffic is managed if it is ever to increase to 3X levels safely.

Despite the high number of separation violations in the manual 3X condition and given that the focus of this study was to explore and compare conflict resolution capabilities across resolution modes, it was interesting to see how the automation provided safety benefits as the traffic levels increased. This was observed as early as the 2X level of traffic where the interactive mode allowed significantly lower numbers of separation violations than the manual mode. This trend continued to an even greater extent at the 3X level of traffic where both the interactive and fully automated resolution modes resulted in measurably fewer separation violations than the manual mode.

The results concerning separation violations suggest that the greatest benefit of the automation would be realized as the levels of traffic increased. However, although the interactive and fully automated resolution modes provided obvious safety benefits, even the results for those conditions would be unacceptable to both industry and the flying public. However, it should be stressed that these results only pertain to a component of the overall Advanced Airspace Concept. The integration with other components for short-term conflict detection and resolution is intended to provide the safety assurance function required to eliminate these separation violations. Perhaps an examination of the concept's remaining components, in addition to the tool used in this study, would yield results that would be more in line with the safety standards in place today.

The minimum separation distance metric was the other measure used to define safety in this study. One interesting item of note that followed the results at the 1X level of traffic was that the time step characteristics shown in Figure 9 were quite similar to the baseline measures used to represent current day traffic in McNally and Gong's 2006 paper from which this metric was borrowed. This was a reassuring observation as it provided support for the use of the 1X scenarios as the foundation for the 2X and 3X levels of traffic.

With respect to the minimum separation distances at the 1X traffic level, the manual and interactive resolution modes were quite comparable to one another. It was the fully automated mode that resulted in the greatest number of aircraft that passed within 10 nm of one another. However, as mentioned earlier, there were no separation violations that occurred in this mode, which is the ultimate measure. Perhaps in this case having to deal with a greater number of aircraft closer in to one another is a price that must be paid in order to avoid a serious incident.

As the traffic levels increased, however, the differences became more apparent between the manual and interactive conditions. At the same time, the interactive and fully automated conditions showed results that were more in line with one another. Particularly at the 3X level of traffic, the manual mode of resolution resulted in the greatest number of aircraft within 10 nm of one another. Contrasted with the automated conditions, the manual mode of resolution not only had the greatest number of separation violations, it also resulted in the most dangerous situation in that so many aircraft were so close to one another that the degrees of freedom to maneuver were severely reduced. This meant that without the automated tool suggesting or executing resolutions at these higher levels of traffic, the path to recovery from a difficult and complex traffic situation would be difficult if not impossible.

Continuing with the 3X level of traffic, another item of interest was the apparent benefit that keeping the aircraft greater than 10 nm from one another provided. As shown in Figure 11, both the interactive and fully automated modes of resolution had much fewer aircraft that were less than 10 nm from each other when compared to the manual mode. Apparently the automated tool provided the ability to keep aircraft outside of 10 nm, which in turn vastly reduced the number of aircraft that ultimately lost separation.

Taken together, the results for the safety aspect of this study showed that separation assurance simply cannot be adequately handled without at least the support of an automated decision support tool. Having the automation available afforded the participants more time and opportunity to deal with the conflicts, which eventually not only significantly cut down on the number of separation violations, but the overall number of aircraft within 10 nm of one another. This not only meant that the immediate traffic situation was safer, but that in the event of an off-nominal situation arising, there would be more space and time with which to address the event than if automation was not available.

Efficiency

Following safety, efficiency was the next concern addressed in this study. Within this larger measure were the three more specific sub-measures of average delay, lateral distance, and vertical distance associated with each resolution clearance. An additional measure included in the analysis of delay was the total delay imposed on the system as a result of all conflict resolutions. Outwardly this measure simply provided a very broad overview of the delay impacts of each resolution mode at the three levels of traffic. This served as a quick and easily quantifiable measure of comparison. From this comparison the results showed that, similar to what was observed in safety, access to the automation in the interactive and fully automated conditions provided benefits relative to the manual condition. In this case it meant that the total delay imposed on the system at the 2X and 3X levels of traffic was consistently less when the automated conflict resolution tool was available. Despite the fact that there were no statistically significant differences between the manual and interactive conditions, Figure 12 shows a clear trend favoring the automation in terms of reduced delay.

One relevant issue that needs to be addressed before proceeding relates to the delay results observed at the 1X level of traffic. The manual mode of resolution in this case actually afforded a time savings. While this might lead one to simply conclude that resolving conflicts manually was more efficient, a potential confound was present that allowed for the greater efficiency. Although the participants were instructed to only send clearances related to conflict resolutions and that their role was limited to that task, some participants were still observed attempting to provide services to both conflict and non-conflict aircraft through extensive route optimization. Although this action was caught and the offending participants ceased their attempts at providing services, the number of these route optimizations was apparently enough to skew the results somewhat at the 1X level of traffic.

Moving past this issue, the results for total delay ultimately reflected more than simply the overall delay. Due to the fact that total delay was a cumulative measure, the results may have been affected by the number of resolutions enacted: the more conflicts that were resolved meant that there could very likely be more overall delay. Interestingly this was not necessarily the case. Although the interactive and fully automated conditions sent more conflict resolutions than the manual condition, particularly at the 3X level of traffic, overall delay was relatively less. This meant that despite sending fewer resolutions to conflict aircraft, the manual mode of conflict resolution resulted in a cumulative amount of delay that exceeded what was observed in the other two modes.

In order to remove the uncertainties brought about by the different number of resolutions and get at a more comparable measure of delay, the average delay per uplink was subsequently analyzed. Results from this analysis made the greater delays associated with the manual condition even more evident than what was seen earlier. By neutralizing the effect of the number of conflict resolutions, it is clear that on a per resolution basis, resolutions in the manual conditions produced much greater delays as the traffic levels increased.

In addition to delay, the average lateral distance that aircraft deviated as part of a resolution was analyzed as part of the overall measure of efficiency. As pointed out in the discussion of delay, results for the manual condition at the 1X level of traffic were somewhat suspect due to the route optimizations that some of the participants were attempting. In this case, the results showed that for the 1X manual condition there was a distance savings associated with each resolution. At the 2X and 3X levels, however, the results for the manual condition were more in line with what might be expected. Interestingly, though, the manual condition did have a lower average than the fully automated condition at the 2X level of traffic. Apparently at this level participants were able to find resolutions on their own that required less deviation than what the automated algorithm could generate.

This ability did not last, however, as the traffic reached the level of 3X. At this point the lateral distances per resolution in the manual condition were quite high relative to the interactive and fully automated conditions suggesting that without having the automated tool available participants had to resort to more extreme lateral deviations in an attempt to ensure adequate separation and maintain some semblance of control. This observation and the statistically significant interaction particularly at the 3X level of traffic once again highlight the important role that automation must play in the accommodation of greater levels of traffic.

The final measure of efficiency analyzed in this study was the average vertical distance associated with each vertical conflict resolution. For this measure, the interesting finding was that unlike the delay and lateral distance measures of efficiency, results for vertical distance showed that as the traffic levels increased the differences between the resolution modes decreased. This was a departure from the previous efficiency measures as the manual mode generally resulted in greater inefficiencies as the traffic level increased.

Another interesting departure observed with this measure was that the average vertical distances associated with each resolution actually decreased as traffic levels increased. This was certainly not the case in the previous efficiency measures where increases in traffic level brought with it overall increases in delay and lateral distance. The reason for this is not entirely clear but one possibility might be that the increased complexity brought about by the increased traffic levels required a more tactical approach to altitude resolutions: whereas altitude resolutions for the 1X and to a lesser extent the 2X level of traffic allowed for greater vertical distances to be used to completely resolve a conflict, the 3X level of traffic may have had fewer safe altitudes with which to work with and may have consequently required participants to solve immediate conflicts with a minimal altitude change in exchange for a secondary conflict downstream that required another tactical, minimal altitude change.

Having discussed the measures of efficiency thus far, one final issue related to the topic concerns the overall reduction in variance observed through the usage of the automated conflict resolution tool. A comparison of the manual and interactive modes for each of the three efficiency measures shows that the standard deviation is nearly always less for the interactive mode, which only became more pronounced at the higher levels of traffic. The reduction of variance and more specifically the introduction of consistency to the task of conflict resolution was, in fact, one of the stated benefits of the automated tool. The results observed for efficiency seem to support that claim and further the case for implementing this type of tool in some capacity in preparation for the looming increase in air traffic.

Workload

As consistency was a stated benefit of the automated conflict resolution tool, the reduction of workload was as well. One of the concerns often cited as being a key barrier to the predicted increases in traffic level is workload. Part of the impetus for the automated conflict resolution algorithm's development was in answer to this concern. Although the ultimate role envisioned for the algorithm and tool that would employ it is in a fully automated capacity, it is assumed that there would be a transitional period where it would serve as a decision support tool for the operators. In this study, the workload of participants was compared between the manual and interactive conditions in order to gauge the automated tool's impact.

Before proceeding any further in the discussion of the workload results, it is important to make a qualification for the measure of workload used in this study. For this study, workload was in reference to the singular task of resolving conflicts. This is in contrast to the wider set of tasks normally performed by controllers such as making and accepting handoffs of aircraft, maintaining voice communications with each aircraft, accommodating arrival and departure traffic for their upcoming phases of flight, as well as the monitoring, detection, and resolution of conflicts. These are but a few of the tasks required of controllers. But even if one were to perform these tasks, a very different picture of workload would emerge. For example, the previous TOOWiLD study (Prevôt et al., 2007) involved participants performing many of the tasks normally required of air traffic controllers as well as some unique tasks at current day levels of traffic, the equivalent of 1X. In this environment, the average workload reported by participants was nearly 4 out of 7. This stands in contrast to the workload ratings in this study at the 1X level of traffic, which rarely exceeded the lowest rating of 1. This difference highlights the fact that workload in this study has a somewhat limited scope and its interpretation and generalization should be treated with care. This does not detract, however, from the more basic issue of whether or not the automation reduces workload.

In this case, the results showed that regardless of whether the average or peak workload ratings were analyzed, the interactive mode of conflict resolution provided a reduction in reported workload at every level of traffic. Trends in the data showed that the automated tool had an increasingly beneficial impact on workload as the traffic level increased. Results from the post-run questionnaires actually showed a significant difference between the two conditions as early as the 2X level of traffic and continued to the 3X level. These results not only provide a basis for support for the proposed benefits of the automated algorithm but it also showed that the benefits in these terms could be reaped early on. In answer to the concerns of workload being a barrier to the increase in traffic, these results highlight the possible instrumental role that this tool could play in the accommodation of the predicted levels of air traffic.

Acceptability

The final concern addressed in this study was the acceptability of the resolutions suggested by the automated conflict resolution algorithm. This was a critical area of interest as the path to implementation of any tool in the air traffic control domain is often dictated by its acceptance from the controllers. To that end, acceptability was measured by the participants' usage characteristics of the automation as well as through a more subjective means via questionnaire.

In terms of how the automation was used in the interactive condition, the obvious trend was that the participants came to rely on the automated resolutions to a far greater extent as the traffic levels increased. Conversely, the percentage of times that those resolutions were modified decreased. This latter measure was used as the proxy for acceptability as the modification of a suggested resolution would imply that it was in some way unacceptable. Taking these results together one could conclude that the acceptability of the automation increased along with traffic levels. Results from the questionnaires somewhat contradict this interpretation, however.

Responses to the questionnaire item showed that the acceptability of the automation's suggested resolution actually decreased as a function of traffic level such that the 3X level of traffic produced significantly lower acceptability ratings than the two lesser levels. That being said, it should be kept in mind that these low ratings were still fairly high with an average response of nearly 6 out of a possible 7. However, the trend of decreasing acceptability cannot be denied, particularly when the trend for the previous measure of acceptability was opposite. This requires further discussion of these results in order to attempt the reconciliation of this disparity.

As mentioned earlier, the participants came to rely on the automation more as the traffic levels increased and accepted its resolutions without modification at a greater rate as well. Given the progressive workload and difficulty associated with the levels of traffic, it seems as though the automation ended up being used almost as a time saving tool. If this was indeed the case, the automation's suggested resolutions may have been more acceptable out of necessity. It is still interesting to see then that the resolutions were subjectively less acceptable given the fact that the resolutions used were, regardless of the measure, almost universally more beneficial than what was observed in conditions without the automation. Perhaps general frustration with the complexity and pace of the 3X level of traffic was in some way included in the ratings of acceptability and could explain the differences in acceptability.

It may have also been that interpreting the participants' usage characteristics of the automation was not necessarily an adequate measure of acceptability. This is not to say that these results were not valuable. It merely points to a possible, yet minor, limitation of the study. A larger limitation worth discussing was the variance observed in many of the results that were presented. The variance between the manual and interactive conditions was so large at times that it very well likely masked some significant results. This was most likely due to the small sample size and the mixture of participant types that were included. This leads to the discussion of some possible directions that future research could take.

Future Research

The first and perhaps most obvious course to take for follow-on research would be to conduct a similar study with a larger, more homogenous sample. However, the results presented from this study beg for further investigation into other relevant issues. One of the first possibilities that stand out is that since this study focused on automation, it naturally follows that an investigation into some of the traditional concerns of automation be addressed. This could involve such issues as trust, bias, and failure recovery. With respect to failure recovery, it would be very interesting to examine how failures in conflict detection would be dealt with in terms of being able to identify that there has been a failure and then having the time and ability to successfully avert a loss of separation. This type of study within the framework of the current study would also allow for the fully-automated conditions to be included in the statistical analysis, which was not the case here. Another possible direction for future research would be to build upon this study by having essentially the same conditions but instead have the participants perform, in addition to resolving conflicts, the other tasks normally associated with the job.

Conclusion

The results from this study highlight two issues that have immediate relevance to the automation analyzed for this study as well as the future viability of the air transportation system in the United States. First, with respect to the automation, the results showed that it provided significant benefits relative to the manual condition in terms of safety, efficiency, and workload. This was particularly true at high levels of traffic. The behavior of the automated conflict resolution tool and underlying algorithm also proved to be generally acceptable. However, these same results highlight the second issue, which is that the NAS and all of its stakeholders face some serious challenges if the predicted threefold increase in air traffic is to be realized. With safety being of paramount importance, the results showed that even with trajectory-based automation, a great deal more must be done fairly quickly in order to be able to maintain safety in the increasingly crowded skies. Despite this fact, the results of this study demonstrated that using the automation for conflict resolution as part of a near term decision support tool was an example of exactly the types of changes that need to take place if the increases in air traffic are to ever be handled safely and efficiently.

REFERENCES

Air Transport Association (2006). *Cost of ATC delays*. Retrieved January 22, 2007, from http://www.airlines.org/economics/specialtopics/ATC+Delay+Cost.htm

Barnett, A. (2001). Air Safety: End of the golden age? *The Journal of the Operational Research Society*, 52(8), 849-854.

Bolczak, R., Gonda, J.C., Saumsiegle, W.J., & Tornese, R.A. (2004). Controller-pilot data link communications (CPDLC) Build 1 value-added services. *Proceedings of the 23rd Digital Avionics Systems Conference (DASC), 2004.*

Bonnefoy, P.A., & Hansman, R.J. (2005). *Emergence of secondary airports and dynamics of regional airport systems in the United States*, (ICAT-2005-02). Massachusetts Institute of Technology, Cambridge, MA.

Carroll, J. R., & Guadiano, N. (2006, December 17). Fewer air controllers could lead to more mistakes, union says. *USA Today*, Retrieved January 18, 2007, from http://www.usatoday.com/travel/flights/2006-12-17-air-traffic_x.htm

Erzberger, H. (2001). The automated airspace concept. *Proceedings of the fourth USA/Europe Air traffic management R&D seminar,* Santa Fe, NM, USA, December 3-7, 2001.

Erzberger, H. (2004). Transforming the NAS: The next generation air traffic control system. *Proceedings of the international congress of the aeronautical sciences (ICAS),*Yokohama, Japan, August 30, 2004.

Erzberger, H. (2006). Automated conflict resolution for air traffic control. In *Proceedings of the25th International congress of the aeronautical sciences (ICAS),* Hamburg, Germany, September 3-8, 2006.

Federal Aviation Administration (2004). *Audit of controls over the reporting of operational errors*(AV-2004-085). Washington DC: Government Printing Office.

Federal Aviation Administration (2004). *A plan for the future: The federal aviation administration's 10-year strategy for the air traffic control workforce.* Washington DC: Government Printing Office.

Federal Aviation Administration (2006). *FAA forecast fact sheet.* Retrieved January 24, 2007 from the Federal Aviation Administration web site: http://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=6886

Federal Aviation Administration Air Traffic Organization (2006). *Moving America safely, 2005 Annual performance report.* Washington DC: Government Printing Office.

Hoekstra, J.M., Ruigrok, R.C.J., & van Gent, R.N.H.W. (2000). Free flight in a crowded airspace. In *Proceedings of the 3rd USA/Europe Air Traffic Management R&D Seminar*, Naples, Italy, June 13-16, 2000.

Jardin, M. (2004). Air traffic conflict models. AIAA 2004-6393, In *Proceedings of AIAA 4th Aviation technology, integration and operations (ATIO) forum*, Chicago, IL, September 20-22, 2004.

Joint Planning and Development Office (2004). *Next Generation Air Transportation System: Integrated plan.* Washington DC: Government Printing Office.

Kirk, D.B, Bowen, K.C., Heagy, W.S., Rozen, N.E., & Viets, K.J. (2001). Development and assessment of problem resolution capabilities for the enroute sector controller. AIAA-2001-5255, In *Proceedings of AIAA aircraft, technology integration, and operations forum,* Los Angeles, CA, October 16-18, 2001.

Kirwan, B., & Flynn, M. (2001). Identification of air traffic controller conflict resolution Strategies for the CORA (Conflict Resolution Assistant) project. In *Proceeddings of the 4th USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, NM, USA, December 3-7, 2001.

Kirwan, B., & Flynn, M. (2002). *Towards a controller-based conflict resolution tool- A literature review.* Technical Report, Eurocontrol, 2002. ASA.01.CORA.2.DEL04-A.LIT.

Krozel, J., Hoffman, B., Penny, S., & Butler, T. (2003). Aggregate statistics of the national airspace system. AIAA-2003-5711. In *Proceedings of AIAA guidance, navigation, and control conference,* Austin, TX, USA, August 11-14, 2003.

Machol, R.E. (1979). Effectiveness of the Air Traffic Control System. *The Journal of the Operational Research Society*, 30(2), 113-119.

Maignan, G. (1992). Safety Aspects of Increased Capacity of Airport and ATS. *Flight Safety Digest*, July, 107-110.

McNally, D., & Gong, C. (2006). Concept and laboratory analysis of trajectory-based automation for separation assurance. In *Proceedings of AIAA guidance, navigation, and control conference and exhibit*, Keystone, CO, USA, August 21-24, 2006.

Mineta, N. Y. (2004, January 27). *Securing America's place as global leader in aviation's second century*. Speech presented for the Aero Club of Washington, Washington DC.

National Air Traffic Controllers Association (n.d.). *Staffing: The workforce plan needs more eyes, too.* Retrieved January 21, 2007, from http://www.natca.org/legislationcenter/staffing.msp

Nolan, M.S. (2004). *Fundamentals of air traffic control* (4th ed.). California: Brooks/Cole-Thomson Learning.

Prevôt, T. (2002). Exploring the many perspectives of distributed air traffic management: The Multi-Aircraft Control System MACS. In S. Chatty, J. Hansman, & G. Boy. (Eds). *Proceedings of the HCI-Aero 2002*, AAAI Press, Menlo Park, CA. pp 149-154.

Prevôt, T., Callantine, T., Homola, J., Lee, P., Mercer, J., Palmer, E., & Smith, N. (2007). Air/ground simulation of trajectory oriented operations with limited delegation. In *Proceedings of the 7th USA/Europe Air Traffic Management R&D Seminar*, Barcelona, Spain, July 2-5, 2007.

APPENDIXES

Appendix A

Post-Run Questionnaire for the Manual Condition

Part-Task Study Jul/Aug 2007
Post-Run Questionnaire

Date:__________ Run#____
Exp. Condition:__M Position:___

Workload

1. **Mental Demand:** What level of mental and cognitive effort (e.g. thinking, analyzing, searching, etc.) did performing the tasks require?

   Low 1 2 3 4 5 6 7 **High**

2. **Effort:** What do you feel your peak level of workload was for the run just completed?

   Low 1 2 3 4 5 6 7 **High**

Operations

3. **Efficiency:** How efficient (e.g. degree of path deviations, magnitude of flight level changes) do you think your clearances were in the resolution of conflicts?

   Low 1 2 3 4 5 6 7 **High**

4. If you feel the clearances could have been more efficient, what factors do you think might have been the cause of the inefficiencies?

   ____
   ____
   ____

5. Rate the level of manageability of the traffic encountered in the completed run.

   Unmanageable 1 2 3 4 5 6 7 **Fully manageable**

Strategies

6. What general strategies did you develop and/or employ as the run progressed with respect to managing the traffic and resolving conflicts?

   ____
   ____
   ____

7. Which conflict resolution method (e.g. lateral path stretch or altitude change), if any, did you come to rely on most and why?

   ____
   ____
   ____

Post-Run Questionnaire for the Interactive Condition

Part-Task Study Jul/Aug 2007 Date:__________ Run#____
Post-Run Questionnaire Exp. Condition:__I Position:___

Workload

8. **Mental Demand:** What level of mental and cognitive effort (e.g. thinking, analyzing, searching, etc.) did performing the tasks require?

Low 1 2 3 4 5 6 7 **High**

9. **Effort:** What do you feel your peak level of workload was for the run just completed?

Low 1 2 3 4 5 6 7 **High**

Operations

10. **Efficiency:** How efficient (e.g. degree of path deviations, magnitude of flight level changes) do you think your clearances were in the resolution of conflicts?

Low 1 2 3 4 5 6 7 **High**

11. If you feel the clearances could have been more efficient, what factors do you think might have been the cause of the inefficiencies?

____

____

12. Rate the level of manageability of the traffic encountered in the completed run.

Unmanageable 1 2 3 4 5 6 7 **Fully manageable**

Automation

13. **Acceptability:** How acceptable do you feel the suggested conflict resolutions from the automation were?

Unacceptable 1 2 3 4 5 6 7 **Acceptable**

14. What were the most common reasons for modifying or rejecting conflict resolutions suggested by the automation?

____

____

Strategies

15. What general strategies, if any, did you develop and/or employ as the run progressed with respect to managing the traffic and resolving conflicts?

____

____

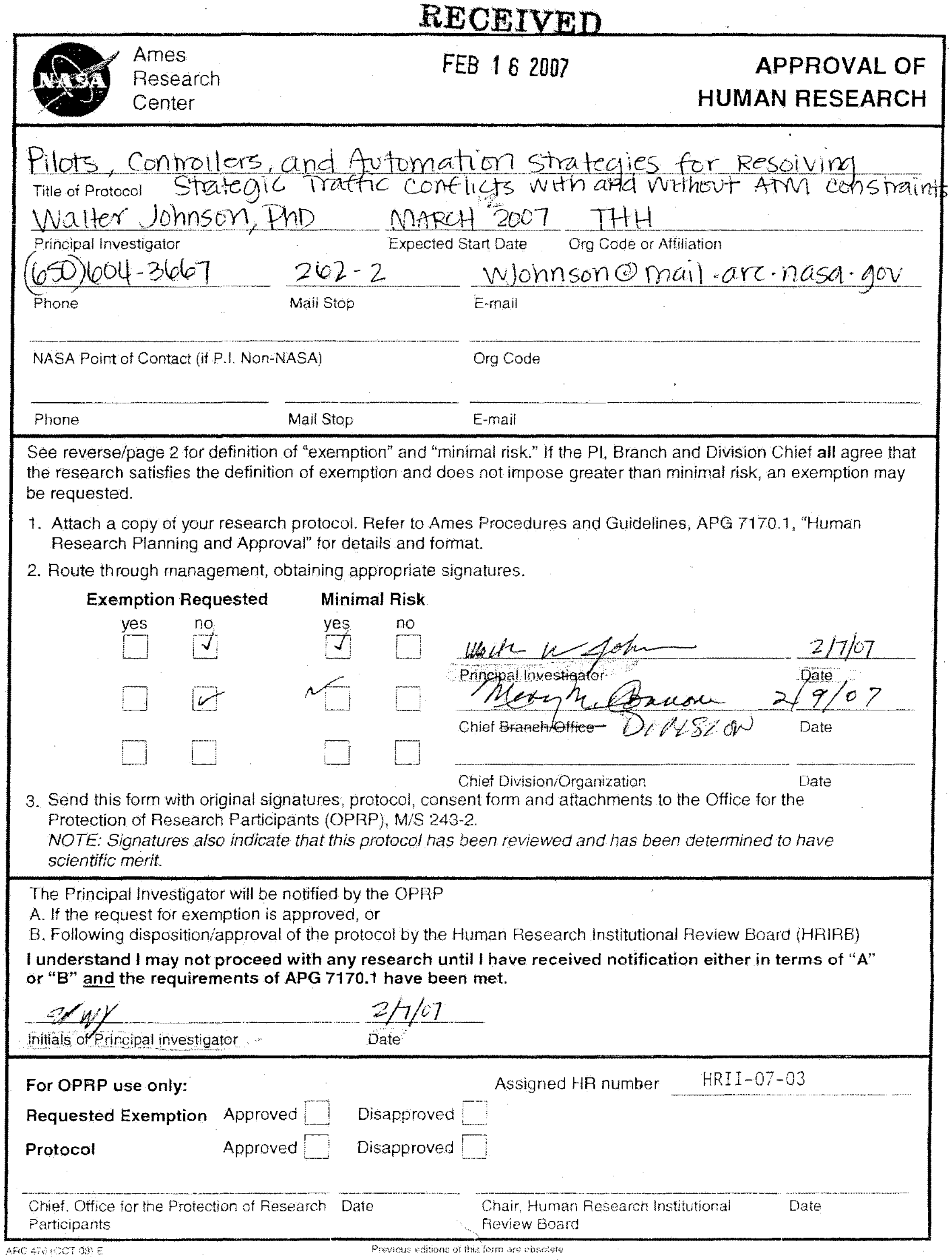
16. Which conflict resolution method (e.g. lateral path stretch or altitude change) did you come to rely on most and why?

____

____

Appendix B

Human Subjects Approval Form



RECEIVED

FEB 16 2007

Ames
Research
Center

APPROVAL OF HUMAN RESEARCH

Title of Protocol: Pilots, Controllers, and Automation Strategies for Resolving Strategic Traffic conflicts with and without ATM constraints

Principal Investigator: Walter Johnson, PhD
Expected Start Date: MARCH 2007
Org Code or Affiliation: THH

Phone: (650)604-3667
Mail Stop: 262-2
E-mail: wjohnson@mail.arc.nasa.gov

NASA Point of Contact (if P.I. Non-NASA):
Org Code:

Phone:
Mail Stop:
E-mail:

See reverse/page 2 for definition of "exemption" and "minimal risk." If the PI, Branch and Division Chief **all** agree that the research satisfies the definition of exemption and does not impose greater than minimal risk, an exemption may be requested.

1. Attach a copy of your research protocol. Refer to Ames Procedures and Guidelines, APG 7170.1, "Human Research Planning and Approval" for details and format.
2. Route through management, obtaining appropriate signatures.

Exemption Requested yes	**Exemption Requested** no	**Minimal Risk** yes	**Minimal Risk** no		
☐	☑	☑	☐	Principal Investigator	Date 2/7/07
☐	☑	☑	☐	Chief ~~Branch/Office~~ DIVISION	Date 2/9/07
☐	☐	☐	☐	Chief Division/Organization	Date

3. Send this form with original signatures, protocol, consent form and attachments to the Office for the Protection of Research Participants (OPRP), M/S 243-2.
NOTE: Signatures also indicate that this protocol has been reviewed and has been determined to have scientific merit.

The Principal Investigator will be notified by the OPRP
A. If the request for exemption is approved, or
B. Following disposition/approval of the protocol by the Human Research Institutional Review Board (HRIRB)

I understand I may not proceed with any research until I have received notification either in terms of "A" or "B" and the requirements of APG 7170.1 have been met.

Initials of Principal Investigator
Date: 2/7/07

For OPRP use only:
Assigned HR number: HRII-07-03

Requested Exemption Approved ☐ Disapproved ☐
Protocol Approved ☐ Disapproved ☐

Chief, Office for the Protection of Research Participants — Date
Chair, Human Research Institutional Review Board — Date

ARC 470 (OCT 03) E

Previous editions of this form are obsolete

IRB Approval

San José State
UNIVERSITY

Office of the Provost
Associate Vice President
Graduate Studies & Research

One Washington Square
San José, CA 95192-0025
Voice: 408-924-2427
Fax: 408-924-2477

E-mail: gradstudies@sjsu.edu
http://www.sjsu.edu

To: Jeffrey Homola

Pamela C Stacks

From: Pamela Stacks, Ph.D.
Associate Vice President
Graduate Studies and Research

Date: March 1, 2007

The Human Subjects-Institutional Review Board has approved your request to use human subjects in the study entitled:

"Analysis of Automated and Air Traffic Control Operator Conflict Resolution Strategies"

This approval is contingent upon the subjects participating in your research project being appropriately protected from risk. This includes the protection of the anonymity of the subjects' identity when they participate in your research project, and with regard to all data that may be collected from the subjects. The approval includes continued monitoring of your research by the Board to assure that the subjects are being adequately and properly protected from such risks. If at any time a subject becomes injured or complains of injury, you must notify Dr. Pamela Stacks, Ph.D. immediately. Injury includes but is not limited to bodily harm, psychological trauma, and release of potentially damaging personal information. This approval for the human subject's portion of your project is in effect for one year, and data collection beyond March 1, 2008 requires an extension request.

Please also be advised that all subjects need to be fully informed and aware that their participation in your research project is voluntary, and that he or she may withdraw from the project at any time. Further, a subject's participation, refusal to participate, or withdrawal will not affect any services that the subject is receiving or will receive at the institution in which the research is being conducted.

If you have any questions, please contact me at (408) 924-2480.

cc. Kevin Jordan, 0120

The California State University:
Chancellor's Office
Bakersfield, Channel Islands, Chico,
Dominguez Hills, East Bay, Fresno,
Fullerton, Humboldt, Long Beach,
Los Angeles, Maritime Academy,
Monterey Bay, Northridge, Pomona,
Sacramento, San Bernardino, San Diego,
San Francisco, San José, San Luis Obispo,
San Marcos, Sonoma, Stanislaus

Appendix C

Consent Form

San José State
UNIVERSITY

Department of Industrial & Systems Engineering
Engr 485
One Washington Square
San Jose, CA 95192-1020
Voice: 408-924-3301
Fax: 408-924-4040
www.engr.sjsu.edu

AGREEMENT TO PARTICIPATE IN RESEARCH AT SAN JOSE STATE UNIVERSITY

Responsible Investigator(s): Jeffrey Homola

Title of Protocol: Analysis of Human and Automated Conflict Resolution Capabilities at Varying Levels of Traffic Density.

You have been asked to participate in a research study investigating the aircraft conflict resolution strategies of air traffic control operators and an automated algorithm.

You will either be asked to manage simulated air traffic at an air traffic control workstation or to fly simulated aircraft at a flight simulation workstation. This study will be conducted at the NASA Ames Research Center's Airspace Operations Laboratory between the dates of March 15, 2007 and June 30, 2007. Your participation will only involve one day with frequent breaks for refreshment and lunch.

There will not be any risks present in this study outside of what are present in daily life.

Direct benefits from participation in this study may include skill maintenance and the gaining of greater insight into the possible advances in the air transportation system. An indirect benefit may be the feeling of reward gained from the knowledge that your participation may be contributing to these advances.

Although the results of this study may be published, no information that could identify you will be included. The data collected from your participation will also be stored on password protected computers, with access granted only to those with the password.

Compensation for your participation will be provided for by Perot Systems based on your qualifications and task.

Questions about this research may be addressed to Jeffrey Homola, (650) 604-4603. Complaints about the research may be presented to Dr. Louis Freund, Ph.D., Department Chair of Industrial & Systems Engineering Department, (408) 924-3890. Questions about a research subjects' rights, or research-related injury may be presented to Pamela Stacks, Ph.D., Associate Vice President, Graduate Studies and Research, at (408) 924-2480.

No service of any kind, to which you are otherwise entitled, will be lost or jeopardized if you choose to "not participate" in the study.

By signing this document, you acknowledge that your consent is being given voluntarily. You may refuse to participate in the entire study or in any part of the study. If you decide to participate in the study, you are free to withdraw at any time without any negative effect on your relations with San Jose State University or with any other participating institutions or agencies.

At the time that you sign this consent form, you will receive a copy of it for your records, signed and dated by the investigator.

Your signature on this document indicates agreement to participate in the study.

The signature of a researcher on this document indicates agreement to include the above named subject in the research and attestation that the subject has been fully informed of his or her rights.

Signature Date

Investigator's Signature Date

The California State University:
Chancellor's Office
Bakersfield, Channel Islands, Chico
Dominguez Hills, East Bay, Fresno,
Fullerton, Humboldt, Long Beach,
Los Angeles, Maritime Academy,
Monterey Bay, Northridge, Pomona
Sacramento, San Bernardino, San
Diego, San Francisco, San Jose, San
Luis Obispo, San Marcos, Sonoma,
Stanislaus