ote technical note techn

The Effect of Voice Communications Latency in High Density, Communications-Intensive Airspace

Randy L. Sollenberger, Ph.D., ACB-220
D. Michael McAnulty, Ph.D., ACB-220
Karol Kerns, Ph.D., AND-300; MITRE/CAASD

January 2003
DOT/FAA/CT/TN03/04

U.S. Department of Transportation
Federal Aviation Administration

William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

Technical Report Documentation Page

1. Report No. DOT/FAA/CT-TN03/04	**2. Government Accession No.**	**3. Recipient's Catalog No.**

4. Title and Subtitle The Effect of Voice Communications Latency in High Density, Communications-Intensive Airspace	**5. Report Date** January 2003
	6. Performing Organization Code ACB-220
7. Author(s) Randy L. Sollenberger, Ph.D., D. Michael McAnulty, Ph.D., ACB-220, and Karol Kerns, Ph.D., AND-300; MITRE/CAASD	**8. Performing Organization Report No.** DOT/FAA/CT-TN03/04
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center NAS Human Factors Group Atlantic City International Airport, NJ 08405	**10. Work Unit No. (TRAIS)**
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address Federal Aviation Administration Air/Ground Communications Product Team 800 Independence Ave., S.W. Washington, DC 20591	**13. Type of Report and Period Covered** Technical Note
	14. Sponsoring Agency Code AND-360

15. Supplementary Notes

16. Abstract

The Federal Aviation Administration (FAA) Next Generation Air-Ground Communications program plans to replace aging analog radio equipment with the Very High Frequency Digital Link Mode 3 (VDL3) system. VDL3 will implement both digital voice and data link communications and will include special features such as controller override, antiblocking, and a transmit status indicator. There are two human factors concerns with the VDL3 system: voice quality and voice throughput delay. Previous research has determined that digital voice technology is highly intelligible and acceptable for Air Traffic Control (ATC) operations. Researchers from the National Airspace System Human Factors Group (ACB-220) of the FAA William J. Hughes Technical Center conducted a high fidelity, human-in-the-loop simulation to examine the impact of voice throughput delay on ATC operations. The communications equipment simulated the VDL3 system with controller override, antiblocking, and transmit status indicator features. The researchers examined ground system delays of 250 ms (current specification), 350 ms (practical alternative), and 750 ms (to demonstrate the sensitivity of the simulation measures) each with their appropriate airborne system delays. Ten controllers from Level 11 and 12 Air Route Traffic Control Centers participated in the study. The results indicated that there were no significant differences between the 250 ms and 350 ms delay conditions. However, the 750 ms condition did produce a significant increase in controller overrides, and the controllers rated it as interfering with some aspects of their communication (e.g., providing optional services). The researchers concluded that the VDL3 system with controller override, antiblocking, and a transmit status indicator can be implemented with a 350 ms ground system delay without causing problems for controllers.

17. Key Words Air Traffic Control, Communications, Digital Voice, Aviation Human Factors, Air Traffic Controller Performance and Workload, Human-in-the-Loop Simulation, Throughput Delay	**18. Distribution Statement** This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161.

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 62	**22. Price**

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Acknowledgments

The following is a final report for Phase I of an operational and human factors evaluation of Very High Frequency (VHF) Digital Link Mode 3 (VDL3) Communications System latency. We based it on a Coordination Draft Plan prepared by Dr. Karol Kerns of the MITRE Center for Advanced Aviation System Development (CAASD). It reflects the contributions and cooperative efforts of a working group composed of representatives of the human factors, requirements, engineering, and union organizations who are responsible for planning and implementing human factors activities associated with the Federal Aviation Administration (FAA) Next Generation Air/Ground Communications (NEXCOM) program. We based the content of this plan on the ideas and issues discussed over many meetings of the NEXCOM Human Factors Working Group.

NEXCOM Human Factors Working Group

Mr. Mark DeBack, NATCA

Ms. Colleen Donovan, AIR-130

Mr. Mike Duer, ANI-90

Mr. Keith Dutch, ATP-410

Ms. Janettarose Greene, ATP-410

Mr. Harry Eberlin, UNITECH

Mr. Tim Henry, ACB-540

Dr. Karol Kerns, MITRE/CAASD

Dr. William Krebs, AAR-100

Mr. Jay Martin, TRIOS

Dr. Michael McAnulty, ACB-220

Mr. Scott McCluskey, ARN-100

Mr. Craig Morris, Thane

Mr. Alan Poston, AND-202

Mr. Rob Strain, MITRE/CAASD

Ms. Terry Stubblefield, AFS-400

Mr. Paul Taylor, PASS

Mr. Paul Travis, AND-360

Mr. Paul Winski, ARN-100

Appreciation is also extended to Dr. Russ Benel, The MITRE Corporation, and Dr. Kim Cardosi, The Volpe National Transportation Systems Center, who reviewed the Coordination Plan and provided their help and expertise in research design, analysis, and human factors. We also thank Dr. Pam Della Rocco from the Human Factors Group (ACB-220) of the Federal Aviation Administration William J. Hughes Technical Center (WJHTC) for providing advice in the research design, analysis, and human factors of this study.

We could not have accomplished this research without the outstanding efforts of the engineering, programming, air traffic subject matter expert, and simulation support staffs of the Research Development and Human Factors Laboratory. Albert Macias led the engineering support staff. Otto Smith was the communications engineer, Mary Delemarre provided system administration and simulation network administration, and George Rowand and Tom D'Ottavi, Jr. engineered the audio-video recording system. Robert Oliver led the computer programming staff. Dario Rossilli and Dan Johnson provided programming support for both developing and and running the simulation. Al Nagy and John Goldman developed the traffic scenarios. Al also participated in shakedown activities and in recruiting the participants. Philip Bassett (Jacksonville Center), Randy Phillips (Cleveland Center), and Doug Kelley (Indianapolis Center) conducted shakedown tests of the scenarios and served as over-the-shoulder observers during the simulations. Mark Hale assisted the Principal Investigator in running the simulation. Kevin Hallman and Fatiha Jackson provided flight progress strip support. Dr. Ferne Friedman-Berg assisted with the data analysis. We thank them all for their contributions to this project. We also thank Samantha Fullerton, who provided runtime support from the Target Generation Facility, the simulation pilots from the Simulation Group, and the participating controllers from Centers across the country.

Table of Contents

Page

Appendices

List of Illustrations

Tables Page

Executive Summary

The Next Generation Air-Ground Communications program plans to replace aging analog radio equipment with a digital system, the Very High Frequency Digital Link Mode 3 (VDL3). VDL3 will implement both digital voice and data link communications and will include special features such as controller override, antiblocking, and a transmit status indicator. There are two human factors concerns with the VDL3 system. The first is the quality of the speech generated by voice coding technology. The second is the effect of additional audio throughput delays introduced by the voice coding and time division multiple access techniques of the VDL3 system configuration. Previous studies by the Federal Aviation Administration William J. Hughes Technical Center (WJHTC) and other International Civil Aviation Organization States have determined that the intelligibility of the digital voice system is acceptable for Air Traffic Control operations. Engineering Research Psychologists from the National Airspace System Human Factors Group (ACB-220) of the WJHTC Research, Development, and Human Factors Laboratory conducted a high fidelity, human-in-the-loop simulation. This study compared system efficiency and controller performance and workload using a simulated VDL3 system with controller override, antiblocking, and transmit status features under three delay conditions based on increasing ground system delays: 250 ms (current specification), 350 ms (practical alternative), and 750 ms (to demonstrate the sensitivity of the simulation measures). Each of the delay conditions also included appropriate delay factors for airborne system processing. The resulting end-to-end delays were somewhat longer for controller-to-pilot than they were for pilot-to-controller transmissions. Ten controllers from the busiest, Level 11 and 12, Air Route Traffic Control Centers participated in the study. The results indicated that there were no significant differences between the 250 and 350 ms delay conditions. However, the 750 ms condition did produce a significant increase in controller overrides, and the controllers rated it as interfering with some aspects of their communication (e.g., providing optional services). The conclusion of the study is that a VDL3 system with antiblocking, controller override and transmit status features and a 350 ms ground system delay would be operationally effective for Air Traffic Service communications and acceptable to controllers.

1. Introduction

The voice communications system currently used by pilots and controllers to exchange Air Traffic Service (ATS) information is functioning near, at, or beyond its planned capacity, and projections foresee continuing increases in the demand for ATS communications for a long time. The Federal Aviation Administration (FAA) Next Generation Air/Ground Communications (NEXCOM) program is developing the Very High Frequency (VHF) Digital Link Mode 3 (VDL3) system to replace the aging analog equipment and provide increased capacity to meet the growing demand for ATS communications. There are two fundamental performance characteristics of the VDL3 technology that are critical determinants of its suitability and acceptability for ATS voice communications in the National Airspace System (NAS): voice quality and voice throughtput delay. Since the late 1980s, researchers have conducted several studies to define and validate the required voice quality of voice coder technology for the VDL3 system (Child, Cleve, & Grable, 1989; Dehel, Grable, & Child, 1989; Farncombe & MacBride, 2000; Fujimori & Ueno, 1999; LaDue, Sollenberger, Belanger, & Heinze, 1997; Renaud, Fistas, Brugere, & Garcia, 1999; Sollenberger, LaDue, Carver, & Heinze, 1997). However, researchers have not examined the impact of the delay characteristics expected in the digital system.

This report describes a high fidelity, controller-in-the-loop simulation study that the FAA will use to evaluate NEXCOM VDL3 delays. In the context of the NEXCOM Human Factors Plan (FAA, 2001), this is one of a series of activities that addresses VDL3 latency performance. The goal of this study was to identify and measure the effects of alternative VDL3 delay parameters in the operational environment. We expected that measures of controller performance, workload, and acceptance would be the most sensitive indicators of delay effects because controllers can compensate for the performance of the air-ground communications system by changing strategies or working harder. In addition, we collected objective and subjective data on system safety, efficiency, and capacity to fully analyze the implications of voice throughput delay for the overall human-system performance.

1.1 VDL3 System Performance and Capabilities

The requirements for the VDL3 air-ground communications system include a maximum limit on voice delay of 250 ms (RTCA, 1994). This 250 ms delay represents the elapsed time from when the controller begins to speak until the audio signal is transmitted through the ground antenna. There is an additional delay of approximately 40 ms until the pilot hears the start of the controller's message (representing the propagation time) and processes the signal through the cockpit avionics. The total end-to-end time from the ground to the air is 290 ms. Air-to-ground communications will have a similar but slightly shorter delay of 260 ms. Without explicit system prompts to indicate channel availability, we also believe that controllers and pilots will continue to follow established procedures when using VDL3 so that waiting to speak and waiting for a reply will be governed by the same timing expectations that apply today. Current analog radio transmissions experience delays between 95 and 150 ms. The current FAA requirements for VDL3 (FAA, 2000) assume that a ground throughput delay of 250 ms will have no adverse effects on Air Traffic Control (ATC) operations.

However, this requirement constrains the design of the VDL3 system. A ground voice delay of 350 ms is more feasible, technically. Consequently, there is considerable interest in conducting further testing to determine whether the specified delay distribution could be relaxed, allowing greater latencies for the average and worst case values.

The VDL3 system requirements also identify new features not available in the current analog system. These features help to compensate for limitations in the current system and to mitigate consequences of throughput delay in the digital system. The VDL3 antiblocking feature actively manages channel access through a controlling ground station that allows only one user to transmit at a time and informs other users who attempt to transmit that the channel is occupied. Another inherent feature of the VDL3 system, controller override or priority access, enables the controller to obtain immediate access to the channel, preempting other users. This feature also allows the controller to free the channel if it is blocked by an active transmitter in a “stuck mic” situation. The transmit status feature provides an indication to pilots when radio transmission is not possible because the channel is busy or because their transmission has been preempted by the controller.

From an operational perspective, the goal for VDL3 voice communications is to remain as consistent as possible with the current voice radio system. However, the consequences of minimizing the voice latency of the VDL3 system are additional design constraints and increased development costs for the future communications system. This research investigated the operational effects of varying latencies on ATS voice communications.

1.2 Effects of Communications Delay on ATC Operations

The temporal characteristics of conversational speech provide important cues that support effective human communication. One important temporal cue is the expected time window for a response. When a response to a message exceeds this expectation time window, the speaker notices a delay, and the quality of the conversation may degrade. In the ATC environment, controllers and pilots have adopted a standard phraseology for conducting spoken dialogues to ensure that communications can be conducted efficiently and with a minimum possibility of error or misunderstanding. ATC communications are also designed with safety measures, such as proper timing and readbacks, to assure that communication is taking place correctly. Given these conversational conventions and rules for controller-pilot interaction, we can assert a number of plausible consequences of delays in operational ATC communications. First, delays may increase the amount of time required to complete each controller-pilot dialogue and the total amount of time devoted to complete required communications tasks. Second, delays may increase the rate of deviations from the standard phraseology and procedures (e.g., partial or missing readbacks) if words or pilot responses are omitted to shorten the dialogues. Third, delays may result in more simultaneous transmissions or retransmissions if the expected time window for a response is exceeded. Finally, delays may result in the untimely delivery of messages as longer transactions are crowded onto a congested communications channel.

Human factors research on the effects of telecommunications delays has identified measures that can be applied to judge the effectiveness of a system. Kitawaki and Itoh (1991) studied delay perception in conversational tasks where subjects experienced delays ranging from 0 to 4 seconds. Subjective assessments were sensitive to increased delays. The subjective assessments included the perceptual, psychological, and behavioral impacts of delay on users. The perceptual measures applied psychophysical constructs to determine delay detection thresholds. The authors assessed psychological impacts via user ratings of satisfaction with delay performances. They assessed behavioral impacts using conversational efficiency scores, indicating the percentage of completion of a task within a certain time interval as compared to the same situation without the delay.

Suganuma (1997) applied these measures in a part-task experimental study to examine the effects of the delays on predefined ATC communications tasks. The results showed that about a quarter of the participating controllers were able to detect a delay increase of 50 ms when it reached 250 ms and half were able to detect it when it reached 450 ms. At the same time, the controller ratings showed that none of the controllers expressed concern or frustration with a delay time of 250 ms. Suganuma also measured a reduction in conversation efficiency with controllers completing fewer communications tasks as delay increased. In the 250 ms delay condition, controllers completed about 95% of the tasks they completed in the 50 ms delay condition.

Farncombe (1997) provides further evidence of ATC communications delay effects in terms of controller communications task performance and flight path efficiency. The author conducted a series of high fidelity, controller-in-the-loop simulations in the extended terminal area airspace to determine whether the VDL3 delay would have an impact on the quality of ATS provided. There were four delay conditions: 130, 280, 400, and 550 ms. The results indicated that the number of transmissions was reduced in the higher delay conditions, but controllers adapted their control strategies by using shorter, reduced messages in some cases and issuing longer, more complex messages in others. Despite the controller's compensation strategy, higher delays resulted in greater variability in the flight paths of arrival traffic, a likely consequence of untimely delivery of instructions. Controllers also reported increased workload and an increased incidence of step-ons during higher delays, although the workload increase was not significant statistically. Finally, the author concluded that the 280 ms delay would not have any adverse effects on operations, but delays greater than 400 ms were unsuitable for the operational environment.

Evidence from Nadler, Mengert, DiSario, Sussman, Grossberg, and Spanier (1993) supports the increased incidence of step-ons and retransmissions as communication workload and throughput delays increased. They conducted a high fidelity, controller-in-the-loop simulation in Atlanta Air Route Traffic Control Center (ARTCC) airspace. They collected data under four delay conditions, 225/0, 169/70, 429/330, and 485/260 ms, representing potential combinations of ground-to-air/air-to ground delays induced by voice switching and satellite transmission equipment, respectively. Longer throughput delays increased the probability that one user would begin to transmit while another user was already sending a message, thus blocking reception of the message in transit. They observed the increased incidence of step-ons only when communications workload was high. Further analyses revealed a significantly higher incidence of step-ons in the 429/330 and 485/260 ms conditions compared to the 225/0 and 169/79 ms conditions. This study also found a corresponding increase in retransmissions. Although the authors expected increased step-ons to result in more missed transmissions (i.e., messages that are blocked and never resent), they did not measure the incidence of missed transmissions.

Related research on ATC data communications (Data Link Benefits Study Team, 1995) provides corroborating evidence on the operational effects attributable to untimely communications. The FAA conducted simulation studies of frequency-congested sectors to show how communications bottlenecks limit system efficiency and to measure the benefits of adding a second data communication channel. The team conducted two experiments in Atlanta ARTCC approach and departure sectors comparing current operations (where flow restrictions and holding arise from saturation of the voice channel) with a future operation (where Controller-Pilot Data Link Communications will be available to augment the air-ground communication channel). The results indicate that an expanded (voice and data) communications channel not only reduced frequency congestion, measured in terms of voice channel occupancy time, it also reduced departure delays, flight times, and distances. In a departure sector, controllers achieved increased throughput because they were able to reliably control departure traffic when in-trail spacing restrictions were relaxed to minimum spacing and to release aircraft to their desired cruise altitudes in a more timely manner. The arrival sector study results were consistent with Farncombe (1997). When the communications channel capacity was increased, controllers provided a more timely service with less vectoring, more direct routes to the arrival fix, and no holding. These data link results do not directly address delays in the air-ground voice communications path, but they indicate that any factor that limits the amount of time available to perform required communications with aircraft may increase controller workload or reduce the timeliness or efficiency with which to accomplish ATC task.

Recently, Rantanen, McCarley, Xu, and Yeakel (2001) designed a study to assess the sensitivity of ATC operations to different magnitudes of audio throughput delay. The study also analyzed the relative contributions of audio throughput delay and delays in pilot response to operational effects. For this study, Rantanen et al. conducted two experiments in which Air Traffic Control Specialists (ATCSs) performed simplified control tasks under four levels of audio delay (150, 250, 350, and 1000 ms) and two levels of pilot response delay (zero or random). They used the zero pilot delay condition to isolate the effect of audio throughput delay, whereas the random pilot delay condition represented realistic system performance. In the random condition, the authors used data from a previous analysis of voice tapes from ARTCCs (Cardosi, 1993) to generate a realistic distribution of response times. The distribution had a mean of 2.5 s and a standard deviation of 2 s that they added to the audio delay. In the first experiment, the controllers made one call to vector traffic. The second experiment required a series of five air-ground communications with each aircraft to resolve a potential conflict, so that they could determine the cumulative effect of audio delays. There was no significant effect of audio delay in the first experiment, and the random pilot delay, which is normally experienced by controllers, actually improved their performance. There were statistically significant effects of increasing audio delay and random pilot delay on total communications duration and on lateral separation in the second experiment. However, the differences between the minimum and maximum delay intervals were very small (3.5 s for communications duration and 0.14 nm for lateral separation). The differences between the three lowest audio delay intervals were inconsistent. The only reliable effect was that performance was generally worse and some elements of workload were slightly higher at the 1000 ms delay than at the other intervals. Neither audio delay nor pilot delay had a significant effect on the number of communications step-ons. In these experiments, the controllers were only communicating with one aircraft at a time, therefore, the study did not address issues about channel access, controller override, or antiblocking.

The previous research provides useful design guidance for the present study. It indicates that although controllers are able to detect communication delays of 250 ms, they can adapt to added delay by changing strategies and altering the timing and content of control instructions. The research further shows that when the delay reaches a certain magnitude, somewhere between 250 and 400 ms, the required compensation techniques may exceed the controllers' capacity or comfort level and translate into more disruptive maneuvers and less efficient flight paths for aircraft. Presently, there are no data on the acceptability of delay parameters between 280 and 400 ms, which includes the technically feasible delay of 350 ms. In this study, we investigated alternative delay parameters to better define an acceptable performance envelope for VDL3.

Because we expected the VDL3 antiblocking feature to mitigate the impact of longer communication delays by minimizing the incidence of step-ons, we incorporated the feature into the present study. We also incorporated the VDL3 controller override and the transmit status indicator features.

The previous research in this area further indicates that these delay effects are context specific and will be observable in a communications-saturated, high workload environment. We selected low altitude en route transition airspace for this study because the traffic characteristics of this environment place rigorous demands on the timing of controller instructions. In addition, busy conditions with high levels of traffic and communications limit the time available for controllers to transmit required instructions.

1.3 Purpose and Hypotheses

The purpose of the study was to evaluate the effects of alternative VDL3 communication delays in a high fidelity, controller-in-the-loop simulation. We examined the impact of alternative delay parameters in terms of

- system safety, capacity, and efficiency;
- air-ground communications; and
- controller workload and acceptance.

We selected three different delay parameters for the study. The first delay was 250 ms, the current specification. The second delay was 350 ms, a feasible and practical alternative. The last delay was 750 ms, which we expected to be unacceptable and demonstrate a negative impact on the system. In general, we expected to observe higher workload and lower controller acceptance and performance with longer communication delays. Specifically, we tested the following hypotheses.

1. Increased communication delay will affect controller task performance and strategies, resulting in fewer, but longer, controller initiated transmissions and increased channel occupancy time.
2. Increased communication delay will result in higher controller workload.
3. Increased communication delay will result in increased controller frustration with air-ground communications and will adversely affect the controller's judgment of system acceptabililty.
4. Increased communication delay will reduce efficiency, resulting in increased flight times and distances.
5. Increased communication delay will adversely affect the perceived margin of system safety.

2. Method

2.1 Participants

Ten male ATCSs from Level 11 and 12 ARTCCs nationwide participated in this study[1]. All participants were nonsupervisory, certified professional controllers who were qualified at their

---

[1] We planned to have 12 participants, but were able to recruit only 11 during the time we ran the simulation. We omitted the data from one of the 11 participants because an equipment problem made the data invalid.

facility and held a current medical certificate. The medical certification ensured that all participants were in good health and had normal or corrected-to-normal vision and hearing. The controllers completed an Informed Consent Form (Appendix A) prior to participating in the study.

Each controller completed a Background Questionnaire (Appendix B) to describe the general demographic characteristics of participants in the study. The controllers ranged in age from 36 to 46 years old ($M = 42.9$, $SD = 3.56$), and ranged in experience from 11 to 26 years of active service ($M = 17.6$, $SD = 4.85$). All participants actively controlled traffic for the past 12 months.

Three Supervisory Air Traffic Control Specialists (SATCSs) served as subject matter experts (SMEs) for the study. The SMEs operated the simulation equipment during shakedown, fine-tuned the traffic flow in scenarios, trained in the use of the observer rating form, and observed the participants during the simulation.

2.2 Test Facility and Equipment

We conducted the simulation in the FAA William J. Hughes Technical Center (WJHTC) Research, Development, and Human Factors Laboratory (RDHFL). Each controller workstation consisted of a high-resolution Sony 2K display, en route keyboard and three-button trackball. The voice communications system consisted of individual relay switchboxes, controller headsets with microphones, and push-to-talk handsets or foot pedals. We provided flight progress strips for each scenario.

The simulation configuration consisted of the Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE) and the Target Generator Facility (TGF). DESIREE emulates Display System Replacement (DSR) functions and receives input from the TGF to display radar targets. In addition to the radar targets, the TGF provides workstations for simulation pilots to communicate with the controllers and enter controller instructions. The TGF controls aircraft maneuvers based on simulation pilot entries and scripted flight plan data.

We set up three controller workstations in one experiment room. The SMEs were positioned behind the controllers to make observations. Engineering Research Psychologists (ERPs) operated the data collection equipment and observed the simulation from an adjacent room. Three simulation pilots supported each controller radar position. The simulation pilot workstations were located in a remote room in the same building. The simulation pilots maneuvered the aircraft using simple keyboard commands and communicated with the controllers using proper ATC phraseology and procedures. Figure 1 presents a schematic diagram (not to scale) showing the interrelationships and communication links used in this simulation.

2.3 Simulation of VDL3 System Latency and Features

We modified the controller-to-pilot communications system to introduce the delay times using a Yamaha D5000 Digital Delay System for each controller position. We simulated three delays for the VDL3 grounds system: 250, 350, and 750 ms. We then added propagation delays and avionics processing times to the simulation resulting in total end-to-end, ground-to-air communication delays of 290, 390, and 790 ms for the three ground system delays, respectively. The corresponding air-to-ground delays for the three delays conditions were 260, 360, and 760 ms, respectively.

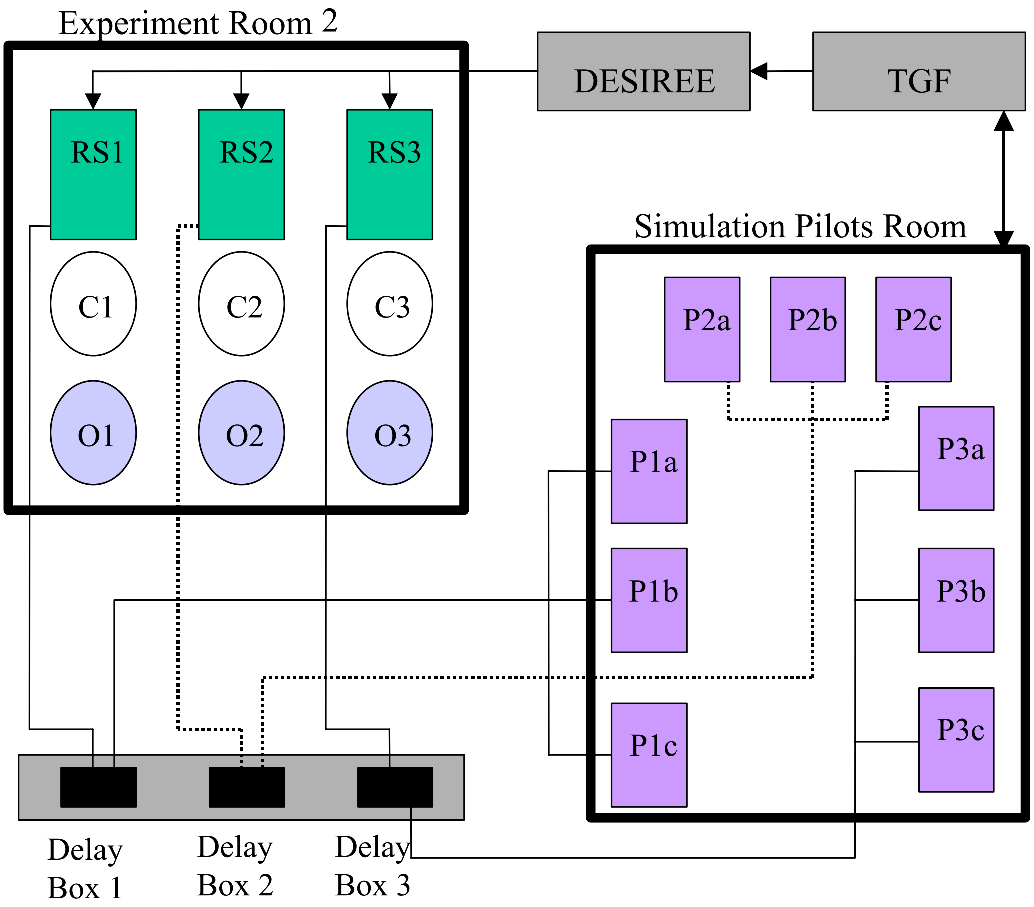


Figure 1. Diagram (not to scale) of the test facility (RS = radar scope, C = controller, O = observer, P = pilot).

We further modified the communications system to implement the controller override, antiblocking, and transmit status indicator features. In our simulation, controller override was automatic and occurred immediately when a controller pressed his handset key or foot pedal[2]. Antiblocking prevented pilot transmissions when the controller or another pilot occupied the communication channel. The transmit status indicator was an audio signal to the pilot that occurred within 500 ms after the pilot attempted a transmission while the communication channel was occupied[3]. Pilots were still able to hear communications on the channel when the transmit status indicator was operating.

---

[2] In the actual VDL3 system, the controller override activation will be slightly delayed based on system latency and the system configuration and timing state.

[3] In the actual VDL3 system, the timing of the onset of the transmit status indicator may be different.

2.4 Airspace

We selected generic en route airspace (Figure 2) for the simulation from a previous human factors study (Yuditsky, Sollenberger, Della Rocco, Friedman-Berg, & Manning, 2002). ERPs and SMEs from the RDHFL designed the generic sector to be a realistic environment for controlling traffic and relatively easy for ATCSs to learn (Guttman & Stein, 1997; Guttman, Stein, & Gromelski, 1995). The sector consisted of "fix" names that they designed to be easily remembered and to simplify operating procedures. The airspace was a low-altitude sector extending from the surface to FL230. The airspace was roughly rectangular in shape and extended for approximately 120 nm from north to south and approximately 85 nm from east to west. The airspace served as a transition sector that fed a terminal area with one major airport and three satellite airports. Arrival routes flowed in a general southbound direction and departure routes flowed in a general northbound direction. The sector contained several intersections and some crossing traffic flow that contributed to sector complexity.

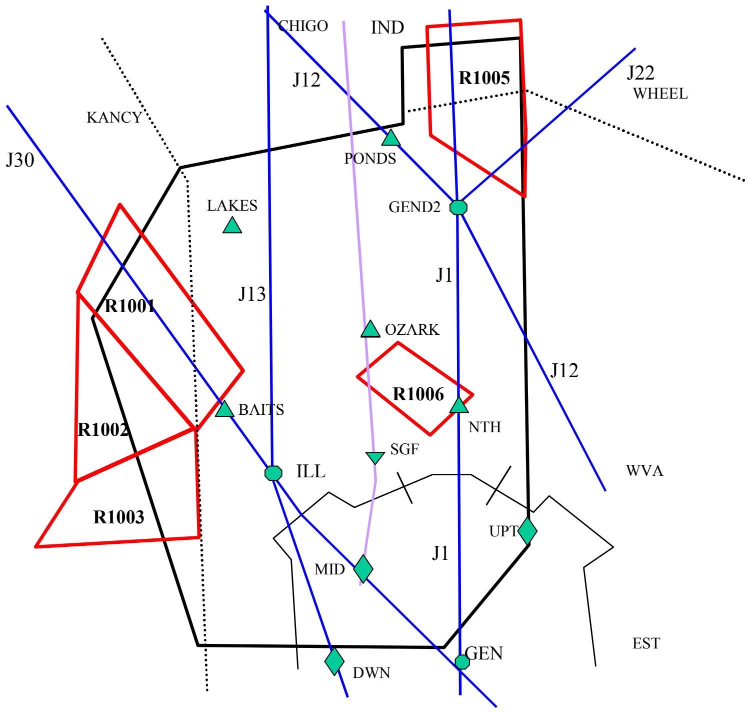


Figure 2. Generic low-altitude en route sector map.

2.5 Traffic Scenarios

We selected traffic scenarios from a previous human factors study (Yuditsky et al., 2002) and modified them to meet the objectives of the present study. We developed three practice scenarios and three test scenarios. The practice and test scenarios were 60 minutes in duration. All test scenarios had the same traffic volume with similar difficulty and complexity; however, aircraft callsigns, spacing, and sequencing were different. Each test scenario consisted of high traffic with 94 total aircraft (54 arrivals, 11 departures, and 29 overflights). The test scenarios had three scripted pilot-initiated transmissions. We used three practice scenarios for training on the generic airspace and the modified communications system. The practice scenarios consisted of moderate traffic with 89 total aircraft (47 arrivals, 15 departures, and 27 overflights). We also developed a warm-up scenario that was 40 minutes in duration and consisted of 88 total aircraft (51 arrivals, 10 departures, and 27 overflights).

2.6 Experimental Design

The study was a single factor, within-subjects (or repeated measures) design with three system delay conditions: 250, 350, and 750 ms. We did not inform the participants or the observers as to which delay condition was in effect for the scenarios. The VDL3 features (controller override, antiblocking, and transmit status indicator) were always operational for all three delay conditions.

Each of the 10 controllers participated in a practice run and three test runs using alternative versions of the test scenarios (A, B, and C) to control for the effects of familiarity with the traffic. We counterbalanced the presentation order of the delays and the alternate versions of the scenarios to minimize sequence effects. Each controller experienced each delay parameter and each version of the scenarios once.

We counterbalanced the delays and alternate versions of the practice scenarios in a similar manner to the test scenarios. Also, we counterbalanced the delay order for the single practice scenario by changing the delays at 10, 20, and 30 minutes into the scenario. The SMEs worked the traffic for the first 10 minutes of the practice scenario. After a handoff briefing, the participants worked the practice scenario for the remaining 30 minutes. Tables 1, 2, and 3 show the presentation order of scenarios and counterbalancing for the study.

Table 1. Presentation Order and Counterbalancing Matrix for Test Scenarios

ATCS	First Problem	Second Problem	Third Problem
1	Scenario A / Delay 750	Scenario C / Delay 250	Scenario B / Delay 350
2	Scenario C / Delay 350	Scenario B / Delay 750	Scenario A / Delay 250
3	Scenario B / Delay 250	Scenario A / Delay 350	Scenario C / Delay 750
4	Scenario B / Delay 750	Scenario C / Delay 350	Scenario A / Delay 250
5	Scenario A / Delay 350	Scenario B / Delay 250	Scenario C / Delay 750
6	Scenario C / Delay 250	Scenario A / Delay 750	Scenario B / Delay 350
7	Scenario C / Delay 750	Scenario B / Delay 250	Scenario A / Delay 350
8	Scenario B / Delay 350	Scenario A / Delay 750	Scenario C / Delay 250
9	Scenario A / Delay 250	Scenario C / Delay 350	Scenario B / Delay 750
10	Scenario B / Delay 350	Scenario C / Delay 250	Scenario A / Delay 750

Table 2. Presentation Order and Counterbalancing Matrix for Practice Scenarios

ATCS	First Problem	Second Problem	Third Problem
1	Practice C / Delay 250	Practice B / Delay 350	Practice A / Delay 750
2	Practice B / Delay 750	Practice A / Delay 250	Practice C / Delay 350
3	Practice A / Delay 350	Practice C / Delay 750	Practice B / Delay 250
4	Practice A / Delay 250	Practice B / Delay 750	Practice C / Delay 350
5	Practice C / Delay 750	Practice A / Delay 350	Practice B / Delay 250
6	Practice B / Delay 350	Practice C / Delay 250	Practice A / Delay 750
7	Practice B / Delay 250	Practice A / Delay 350	Practice C / Delay 750
8	Practice A / Delay 750	Practice C / Delay 250	Practice B / Delay 350
9	Practice C / Delay 350	Practice B / Delay 750	Practice A / Delay 250
10	Practice C / Delay 750	Practice A / Delay 350	Practice B / Delay 250

Table 3. Counterbalancing Matrix for Warm-up Scenario

ATCS	10 Minutes	20 Minutes	30 Minutes
1	Delay 350	Delay 750	Delay 250
2	Delay 250	Delay 350	Delay 750
3	Delay 750	Delay 250	Delay 350
4	Delay 350	Delay 250	Delay 750
5	Delay 250	Delay 750	Delay 350
6	Delay 750	Delay 350	Delay 250
7	Delay 350	Delay 750	Delay 250
8	Delay 250	Delay 350	Delay 750
9	Delay 750	Delay 250	Delay 350
10	Delay 250	Delay 750	Delay 350

2.7 Procedure

The participants arrived at the RDHFL in groups of three controllers for each 2-day simulation session[4]. Each participant worked independent traffic scenarios that did not involve handoffs with the other participants. The ATC simulator automated adjacent sector functionality. On the

---

[4] One group had only two participants.

first day of the study, we briefed the participants on the project goals and the generic airspace. In the afternoon, the controllers received airspace training by working three, 60-minute, practice scenarios, representing each of the delay conditions. A question and answer period followed the hands-on training at the end of the day. On the second day, the controllers worked a practice scenario followed by three, 60-minute, test scenarios. At the end of the day, we conducted a final group debriefing to discuss the participants' experience with the communication delays in the simulation. The participants worked from 8:00 AM to 4:30 PM each day with a rest break after each scenario and a lunch break. Table 4 shows the participant schedule for each simulation session.

Table 4. Participant Simulation Session Schedule

Day 1		**Day 2**	
Time	**Activity**	**Time**	**Activity**
8:00 - 10:00	Project Briefing, Airspace Overview, and Initial Forms	8:00 - 8:40	Warm-up Scenario
		8:40 - 9:00	Break
		9:00 - 10:00	Test Scenario
10:00 - 10:30	Break	10:00 - 10:30	Break
10:30 - 11:30	Practice Scenario	10:30 - 11:30	Test Scenario
11:30 - 1:00	Lunch Break	11:30 - 1:00	Lunch Break
1:00 - 2:00	Practice Scenario	1:00 - 2:00	Test Scenario
2:00 - 2:30	Break	2:00 - 2:30	Break
2:30 - 3:30	Practice Scenario	2:30 - 4:30	Final Debriefing
3:30 - 4:00	Break		
4:00 - 4:30	Question & Answer		

On the first day of the study, the participants signed an Informed Consent Form (Appendix A) and completed a Background Questionnaire (Appendix B). During the practice scenarios, we familiarized the participants with a battery of questionnaires and data collection procedures designed to evaluate the impact of the communication delays. The participants completed a Post-Scenario Questionnaire (Appendix C) and a Controller Acceptance Rating Scale (CARS) form (Appendix D) after each test scenario to assess the impact of the alternative delay conditions. The SMEs completed an Observer Rating Form (Appendix E) after each test scenario. On the last day of the study, the participants completed an Exit Questionnaire (Appendix F).

We used the Air Traffic Workload Input Technique (ATWIT) to assess controller workload during the scenario. ATWIT provides an unobtrusive and reliable means for collecting self-report workload ratings as controllers work traffic (Stein, 1985, 1991). A laptop computer with connected 10-button keypads collected and recorded participant responses. The participants indicated their current workload by pressing one of the keypad buttons labeled from 1 (indicating low workload) to 10 (indicating high workload). The system prompted participants for input every 5 minutes by emitting several beeps and lighting the buttons on the keypad. The participants had 20 seconds to respond by pressing one of the 10 buttons. If participants were too busy to respond within the allowed time, the system recorded a workload rating of 10 by default. However, we excluded these default ratings from further analyses.

2.8 Measures

We collected a battery of objective and subjective measures to assess the impact of the alternative delay conditions. In addition, audio-visual equipment recorded the controllers' actions and communications during the simulation in case we wished to review the simulation at a later date.

2.8.1 Air Traffic Measures

The TGF recorded the position and status of all aircraft every second during the scenarios. We processed these raw data using a data reduction and analysis tool to produce the following standard set of ATC simulation measures (Buckley, DeBaryshe, Hitchner, & Kohn, 1983; Stein & Buckley, 1992).

- Total flight time and flight distance per aircraft
- Number of altitude, heading, and airspeed changes
- Number of loss-of-separation errors

2.8.2 Communication Measures

The RDHFL communications system recorded the time of each push-to-talk key press and release for both controllers and pilots. We processed these raw data using a communications analysis program that computed the following measures.

- Number of controller transmissions, cumulative duration, and average duration of transmissions
- Number of pilot transmissions, cumulative duration, and average duration of transmissions
- Number of controller override transmissions defined as controller transmissions that were initiated while a pilot occupied the communication channel
- Number of pilot blocked transmissions defined as attempted pilot transmissions that were initiated while the controller or another pilot occupied the communication channel
- Mean pilot response interval defined as the time from controller key release to pilot key press

2.8.3 Controller Workload Measures

The controllers provided workload ratings using two different techniques. The first technique was ATWIT, a real-time, unidimensional workload rating method (Stein, 1985, 1991). The second technique was a modified form of the National Aeronautics and Space Administration (NASA) Taskload Index (TLX), a post-scenario, multidimensional workload rating method (Hart & Staveland, 1987). The NASA TLX includes six different subscale ratings (included in the Post-Scenario Questionnaire, Appendix C). The controllers also provided communications workload ratings after each scenario.

2.8.4 Controller Subjective Ratings

The participants rated their performance, situation awareness, pilot performance, and scenario difficulty after each scenario. The participants also evaluated the impact of the communication delays by providing the following ratings after each scenario (Appendix C).

- Extent delay interfered with communications
- Extent delay interfered with control strategy
- Extent delay interfered with timing of control instructions
- Extent delay interfered with providing optional ATC services
- Extent delay interfered with using correct ATC phraseology
- Extent delay interfered with speech clarity
- Extent delay increased speech rate

In addition, the participants rated the realism of the simulation equipment, airspace, scenarios, and overall experience at the end of the study (Appendix F).

2.8.5 Observer Subjective Ratings

The SMEs observed the controllers and made over-the-shoulder ratings during each scenario. The SMEs used an observation form specially designed for ATC simulation observations (Sollenberger, Stein, & Gromelski, 1996; Vardaman & Stein, 1998). The form includes six major rating scales and several subscales for a total of 27 different observer ratings (Appendix E). The SMEs also rated the frequency of the following key events.

- Failed to climb departing aircraft (i.e., forced to level)
- Failed to descend arriving aircraft
- Failed to accept handoff in a timely manner
- Failed to initiate handoff in a timely manner
- Failed to switch frequency in a timely manner
- Failed to issue clearance at appropriate time (i.e., too late or too early)
- Failed to comply with Letters of Agreement

In addition, the SMEs provided an overall operational assessment of ATC safety after each scenario.

2.8.6 Controller Operational Acceptability

We used CARS to measure how well the controllers and the system performed during each scenario (Lee, Kerns, & Bone, 2001). Using CARS, the controllers rated the joint controller-system performance on a numeric scale indicating the degree of problems or deficiencies experienced in each scenario (Appendix D).

3. Results

We used a univariate approach to analyzing the experimental data. We performed a one-way analysis of variance with repeated measures on the delay factor for each of the simulation measures. In each analysis, the standard significance level was $p < .05$. When the delay effect was statistically significant, we conducted Tukey HSD post hoc comparisons at $p < .05$ to determine which of the three means was different from the others.

In this study, we are primarily concerned with determining if longer transmission delay intervals have any negative impact on ATCSs and system efficiency and safety. If there is any possibility of a significant effect, we wanted to detect it, even though we had a relatively small number of participants. Therefore, we have chosen to retain the standard criterion level for all significance tests rather than making changes to the criterion or analytic approach that could lead to erroneous conclusions that there were no negative impacts when they do exist.[5]

When conducting studies with applied relevance, frequently evaluating complex behavior with small samples, Wickens (1998) argues that the primary goal is to establish which conditions are better, by how much, and how confident we are (Wickens). Following his recommendations, we report the alpha levels corresponding to the obtained probability to provide information about statistical confidence. In the following subsections, we also present the results of the study as a series of graphs depicting the mean for each delay condition with an error bar. The error bar represents +/- 1 standard error of the mean as an indicator of between-subject variability. On the graphs, it also indicates the extent of overlap in the data obtained under the different delay conditions. We used the following formula for computing the standard error.

Standard Error = Standard Deviation / Square Root (N), where N is the number of observations used to compute the mean, (i.e., 10 participants)

3.1 Air Traffic Measures

Figure 3 shows the mean flight time and flight distance per aircraft for each of the delay conditions. On average, each aircraft flew for approximately 10.5 minutes and approximately 62 nm through the sector. There were no significant differences between the delay conditions for either measure.

---

[5] Establishing a criterion of $p < .05$ for statistical significance means that 5 out of 100 such test results may have occurred by chance rather than there being a real effect of the independent variable on the variable being measured. The researcher selects the criterion level, and .05 is the most common level chosen. Controller performance is very complex, so we measured many variables to try to evaluate any possible effect of the delay interval. Conducting multiple tests may increase the probability of an erroneous conclusion above the stated probability level. There are numerous statistical techniques for addressing this concern, such as setting a lower criterion level, adjusting the level based on the number of tests, and conducting multivariate tests. All of these techniques make the test more conservative, and thus less likely to detect a real difference when it exists. This is the concept of the power of the test, the probability that it will result in a significant result when it does exist, which is highly influenced by the number of participants. The fewer the number of participants, the less likely that the statistical test will be significant, even if the effect is real.

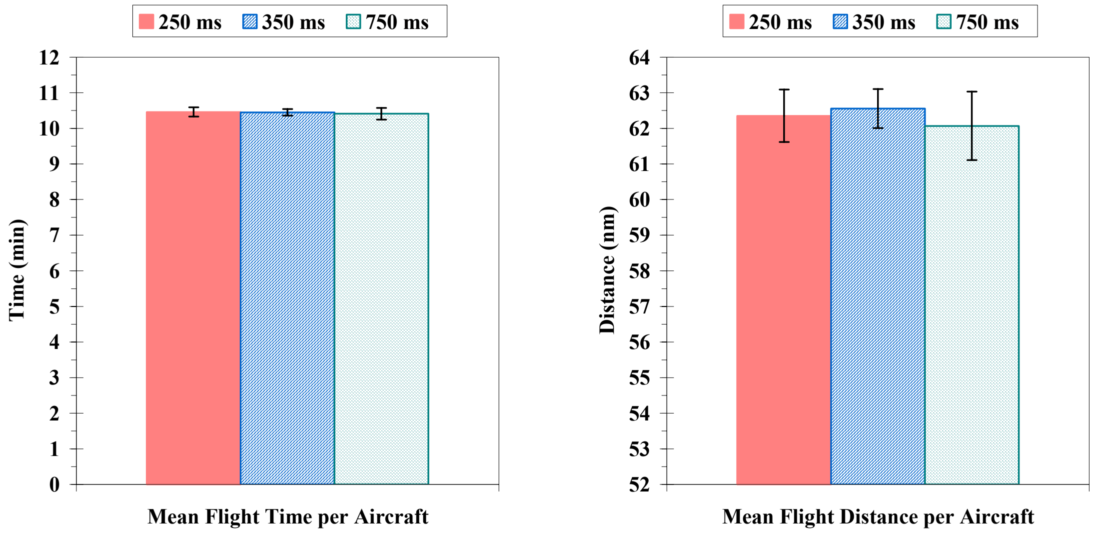


Figure 3. Mean flight time and flight distance per aircraft by delay condition.

Figure 4 shows the mean number of altitude, heading, and airspeed changes for each of the delay conditions. Controllers used altitude changes much more frequently than heading and airspeed changes. On average, there was slightly more than one altitude change per aircraft, but controllers only gave about 12% - 29% of the aircraft heading or airspeed changes. There were no significant differences between the delay conditions for any of the measures.

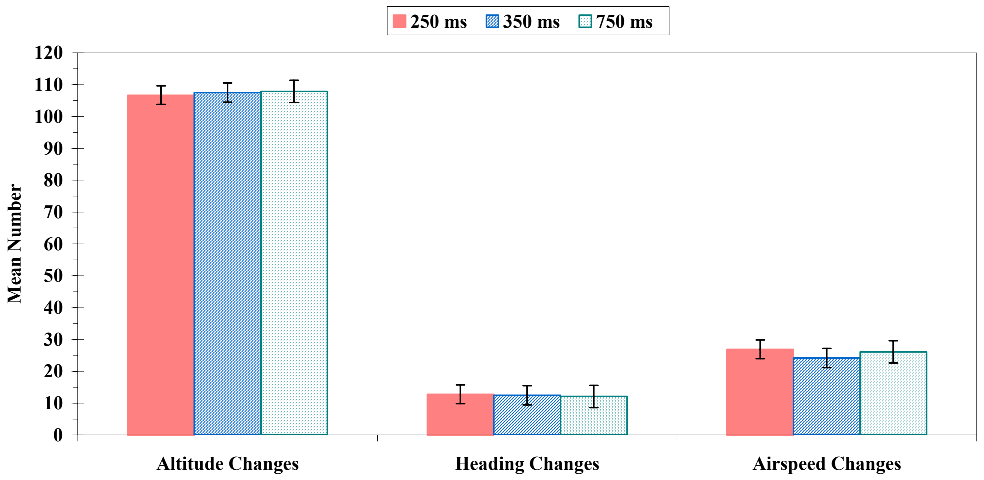
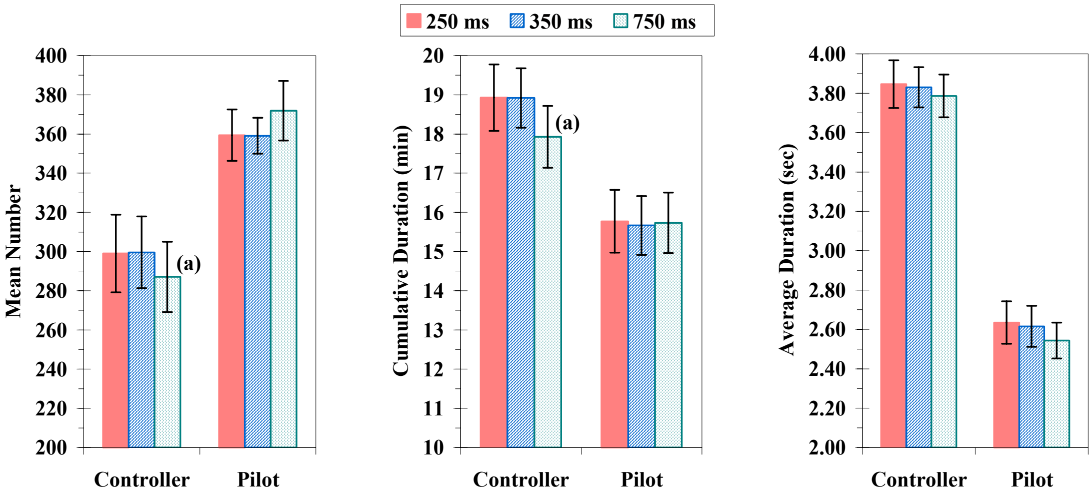


Figure 4. Mean number of altitude, heading, and airspeed changes by delay condition.

3.2 Communication Measures

Figure 5 shows the mean number of controller and pilot transmissions, cumulative duration of transmissions, and average duration per transmission for each of the delay conditions. Controllers used fewer but longer transmissions than pilots and occupied the communication channel longer than pilots. There were two statistically significant delay effects. First, controllers made fewer transmissions during the 750 ms delay compared to the 250 ms and 350 ms delays [$F(2,18) = 5.05$, $p = .018$]. On average, the controllers made 287 transmissions in the 750 ms delay condition compared to approximately 299 transmissions in the other two delay conditions, a decrease of about 4%. Also, controllers occupied the communication channel for a shorter duration during the 750 ms delay compared to the 250 ms and 350 ms delays [$F(2,18) = 9.99$, $p = .001$]. On average, the controllers occupied the communication channel for 18 minutes in the 750 ms delay condition compared to approximately 19 minutes in the other two delay conditions, a decrease of about 5%. Because the number of heading, altitude, and speed clearances and the number of aircraft entering and exiting the sector were essentially equal in all delay conditions, it is likely that controllers did not provide optional services as frequently under the 750 ms condition, which may account for the significant reduction in controller transmissions.



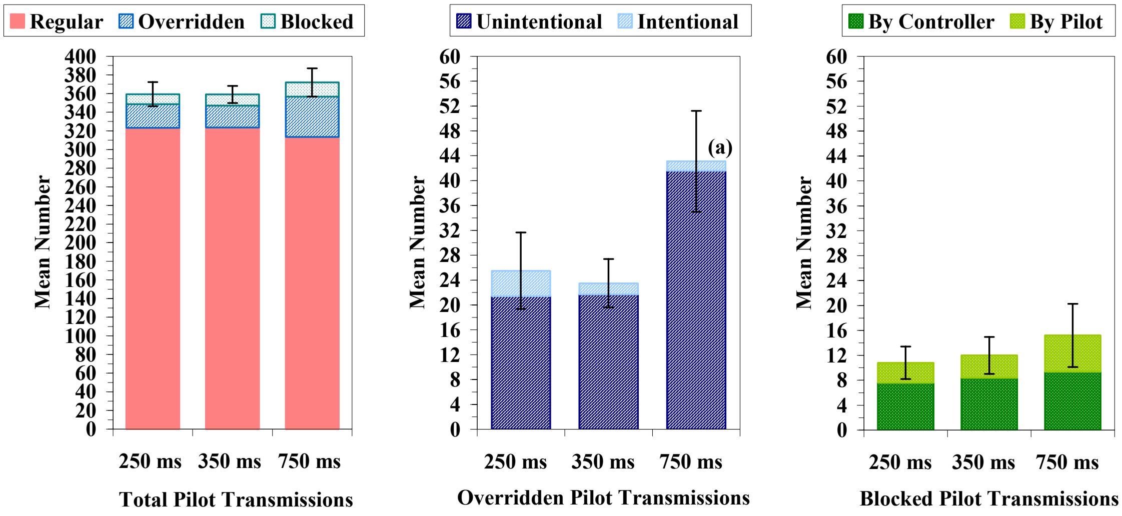
(a) 750 ms < 250 ms & 350 ms, ($p < .05$)

Figure 5. Mean number of controller and pilot transmissions, cumulative duration of transmissions, and average duration of each transmission by delay condition.

Figure 6 depicts the mean number of pilot transmissions and the number of overridden and blocked transmissions for each of the delay conditions. We operationally defined controller overrides as controller transmissions that were initiated while a pilot occupied the communication channel. We defined blocked transmissions as attempted pilot transmissions that were initiated while the controller or another pilot occupied the communication channel. Overridden and blocked transmissions represented a small percentage of total pilot transmissions. Overridden transmissions were only 7%, 7%, and 12% of all pilot transmissions for the 250, 350, and 750 ms delays, respectively. Blocked transmissions were only 3%, 3%, and 4% of all pilot transmissions for the 250, 350, and 750 ms delays, respectively.

Figure 6 also shows the mean number of overridden pilot transmissions divided into transmissions unintentionally overridden by controllers and transmissions intentionally overridden by controllers. We operationally defined an unintentional override as a controller transmission that was initiated during the delay time before the controller began to hear the pilot transmission. Most of the overridden pilot transmissions were unintentional by controllers. Unintentionally overridden transmissions were 84%, 92%, and 97% of all overridden pilot transmissions for the 250, 350, and 750 ms delays, respectively. The number of pilot transmissions overridden by controllers (unintentional and intentional combined) was significantly greater during the 750 ms delay compared to the 250 ms and 350 ms delays, [$F(2,18) = 16.42$, $p = .000088$].

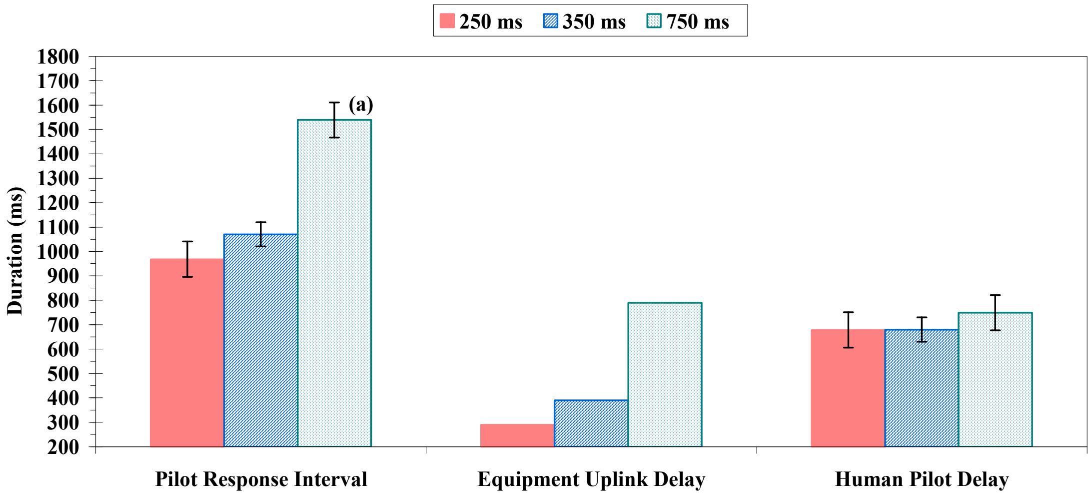
Finally, Figure 6 shows the mean number of blocked pilot transmissions divided into transmissions blocked by controllers and transmissions blocked by another pilot. Controllers blocked about twice as many of the blocked pilot transmissions. Pilot transmissions blocked by controllers were 70%, 70%, and 62% of all blocked pilot transmissions for the 250, 350, and 750 ms delays, respectively. There were no significant differences between the delay conditions for blocked pilot transmissions.



(a) 750 ms > 250 ms & 350 ms, ($p < .05$)

Figure 6. Mean number of regular, overridden, and blocked pilot transmissions by delay condition.

Figure 7 shows the mean pilot response interval, which consisted of an equipment uplink delay and a human pilot delay component, for each of the delay conditions. We operationally defined the pilot response interval from the push-to-talk data as the time from controller key release to pilot key press. We computed the human pilot component by subtracting the equipment uplink delay from the pilot response interval. The pilot response interval statistically increased during the 750 ms delay compared to the 250 ms and 350 ms delays, [$F(2,18) = 68.72$, $p < .000001$]. However, simulation pilots showed a consistent human delay across communication delays. The increasing pilot response interval was due to the equipment uplink delay setting.



(a) 750 ms > 250 ms & 350 ms, (p < .05)

Figure 7. Mean pilot response interval, equipment uplink delay, and human pilot delay-by-condition.

3.3 Controller Workload Measures

Figure 8 shows the mean ATWIT workload ratings with the number of aircraft handled for each 5-minute interval of the scenarios and the overall mean workload ratings for each of the delay conditions. The number of aircraft handled was nearly identical for each of the delay conditions, therefore, the data were collapsed across delays and plotted as one curve. Controllers responded to 97.5% of the ATWIT probes and allowed only 2.5% to time out. We treated the probes to which participants did not respond as missing data for the analyses. Controller mean ATWIT workload ratings ranged from 3.56 to 7.10 on a 10-point scale. In general, controller workload increased with an increasing number of aircraft handled. Workload was lowest at the start of scenarios when the number of aircraft handled was low. Workload was highest at two times in the scenarios when the number of aircraft handled peaked. On average, the ATWIT workload rating was 5.47, which indicates a moderate workload level. There were no significant differences between the delay conditions for ATWIT workload ratings.

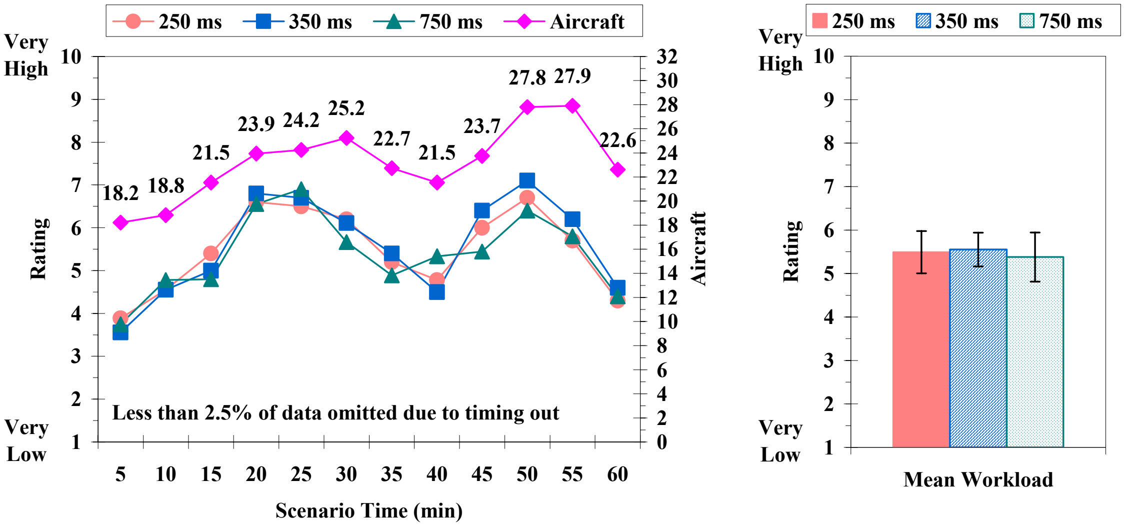


Figure 8. Mean ATWIT workload ratings with number of aircraft handled during scenarios and overall mean workload ratings by delay condition.

Figure 9 shows the mean NASA TLX subscale workload ratings for each of the delay conditions. Controller mean NASA TLX workload ratings ranged from 4.20 to 8.10 on a 10-point scale. In general, ratings were the lowest for the frustration subscale and highest for the performance and effort subscales. There were no significant differences between the delay conditions for any of the NASA TLX workload subscale ratings.

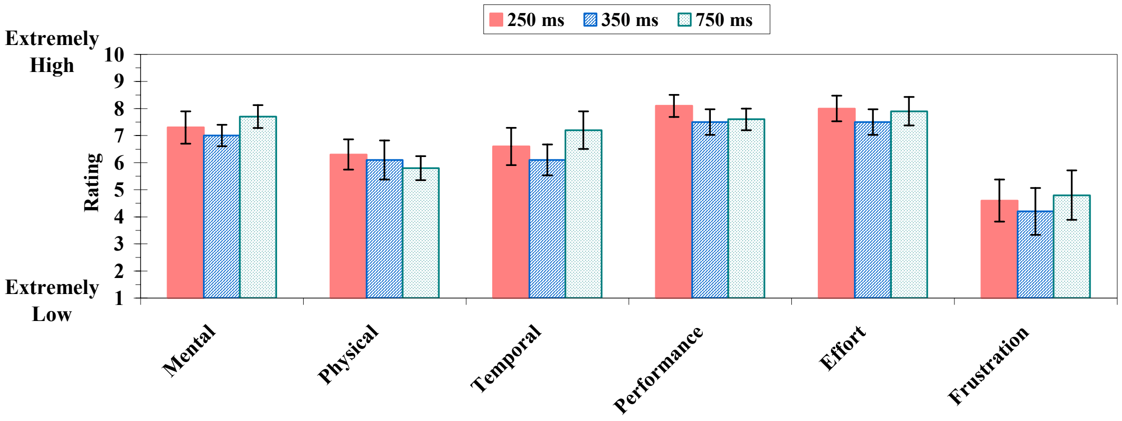


Figure 9. Mean NASA TLX workload ratings for each subscale by delay condition.

3.4 Controller Subjective Ratings

Figure 10 shows controller mean ratings of performance, situation awareness, communications workload, pilot performance, and scenario difficulty for each of the delay conditions. The mean ratings were high and ranged from 7.20 to 8.70 on a 10-point scale. There were no significant differences between the delay conditions for any of the ratings.

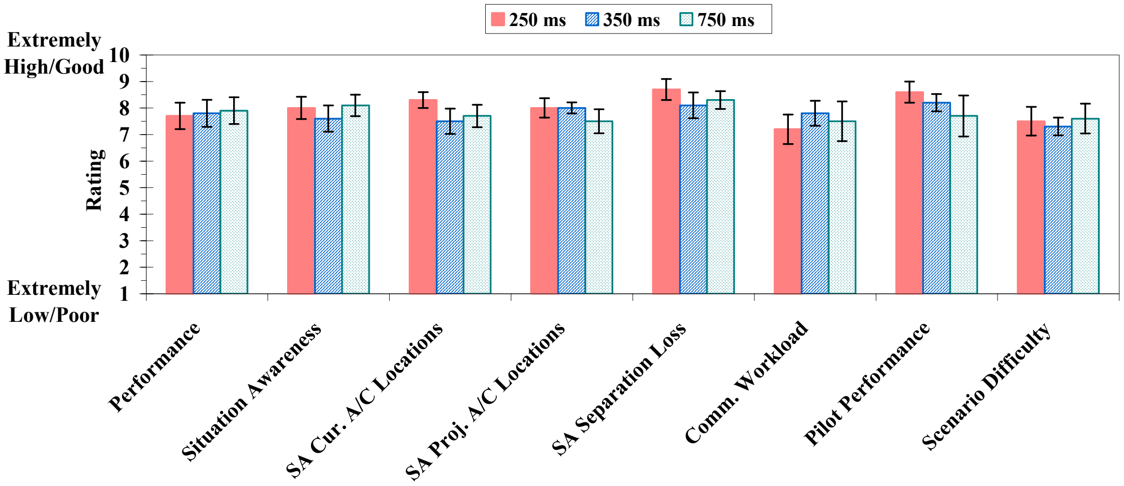


Figure 10. Mean performance, situation awareness, communications workload, pilot performance, and scenario difficulty ratings by delay condition.

Figure 11 shows controller mean ratings of the impact of communication delays for each of the delay conditions. Controllers rated the impact of communication delays in terms of the degree of interference with communications, control strategy, control instructions, optional ATC services, correct phraseology, speech clarity, and degree of increase in speech rate. The mean ratings were low to moderate and ranged from 1.90 to 5.70 on a 10-point scale. There were three statistically significant delay effects. First, controller ratings of the degree of interference with communications were higher for the 750 ms delay compared to 350 ms delay [F(2,18) = 4.14, *p* = .033]. On average, the controllers rated the 750 ms delay as 1.5 points (26%) higher than the 350 ms delay. Second, controller ratings of the degree of interference with speech clarity were also higher for the 750 ms delay compared to 350 ms delay [*F*(2,18) = 4.14, *p* = .033]. On average, the controllers rated the 750 ms delay as 1.4 points (42%) higher than the 350 ms delay. Finally, controller ratings of the degree of interference with optional ATC services were higher for the 750 ms delay compared to both the 250 ms and 350 ms delays [*F*(2,18) = 6.52, *p* = .007]. On average, the controllers rated the 750 ms delay as approximately 1.2 points (28%) higher than the 250 ms and 350 ms delays.

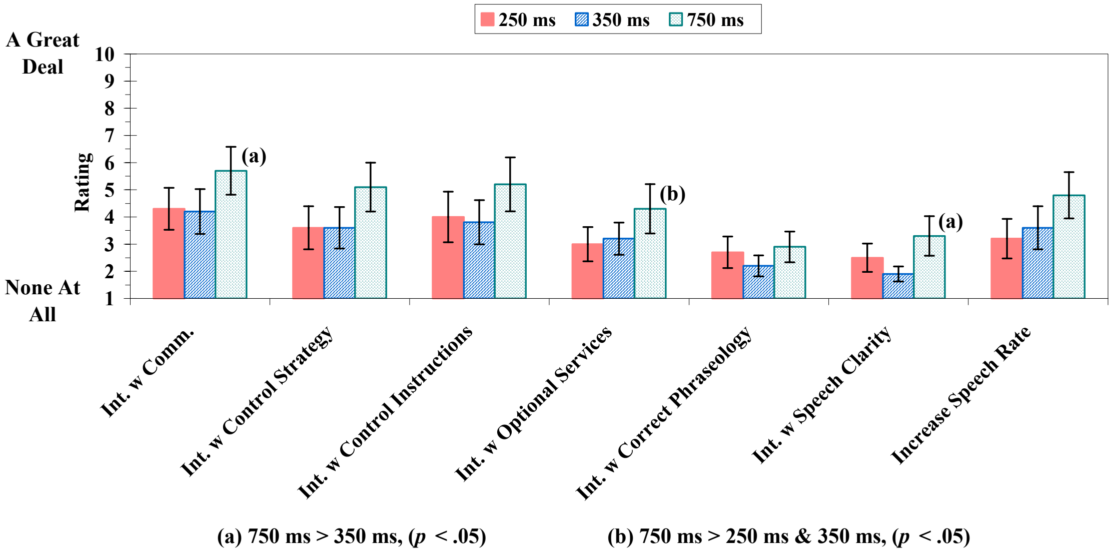


Figure 11. Mean impact of communication delays ratings by delay condition.

3.5 Observer Subjective Ratings

Figure 12 shows observer mean ratings of controller performance on the major scales of the Observer Rating Form for each of the delay conditions. The mean ratings were highly to very highly effective and ranged from 6.30 to 7.40 on an 8-point scale. There were no significant differences between the delay conditions for any of the major scale ratings.

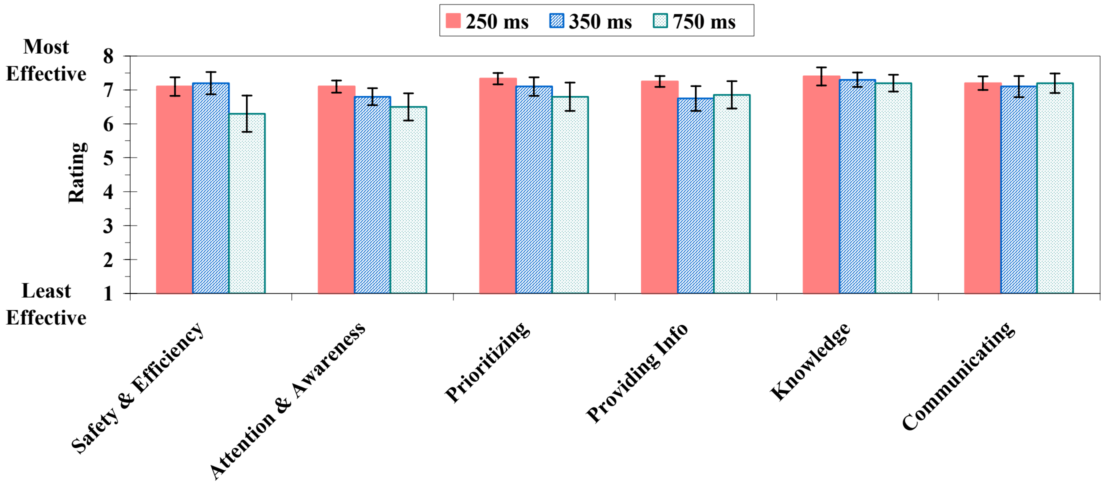


Figure 12. Mean observer ratings of controller performance for each major scale by delay condition.

Figure 13 shows observer mean ratings of controller performance on the subscales of the Safety and Efficiency category for each of the delay conditions. The mean ratings were highly to very highly effective and ranged from 6.10 to 7.20 on an 8-point scale. There were no significant differences between the delay conditions for any of the subscale ratings.

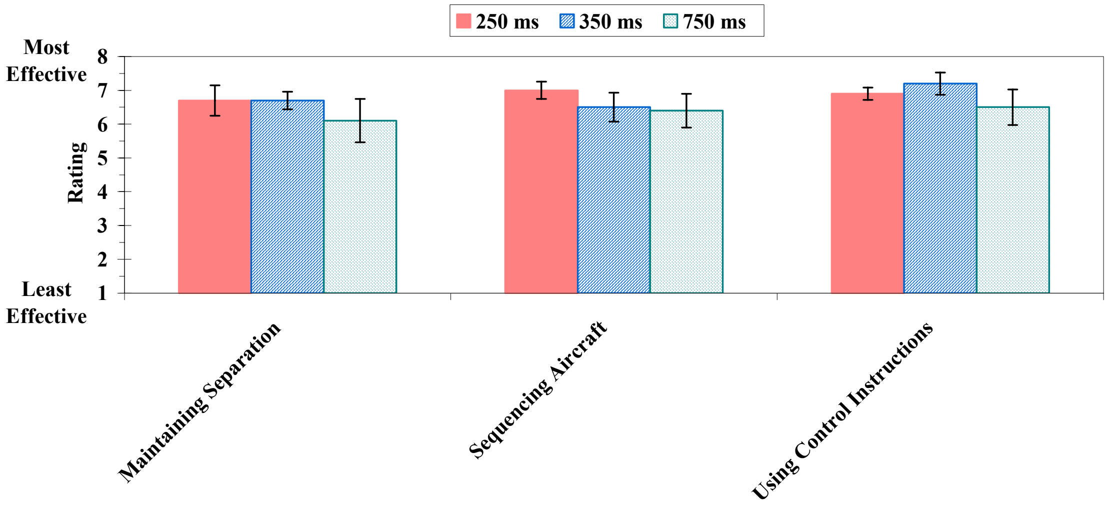


Figure 13. Mean observer ratings of controller performance for each Safety and Efficiency subscale by delay condition.

Figure 14 shows observer mean ratings of controller performance on the subscales of the Attention and Awareness category for each of the delay conditions. The mean ratings were highly to very highly effective and ranged from 6.00 to 7.50 on an 8-point scale. There were no significant differences between the delay conditions for any of the subscale ratings.

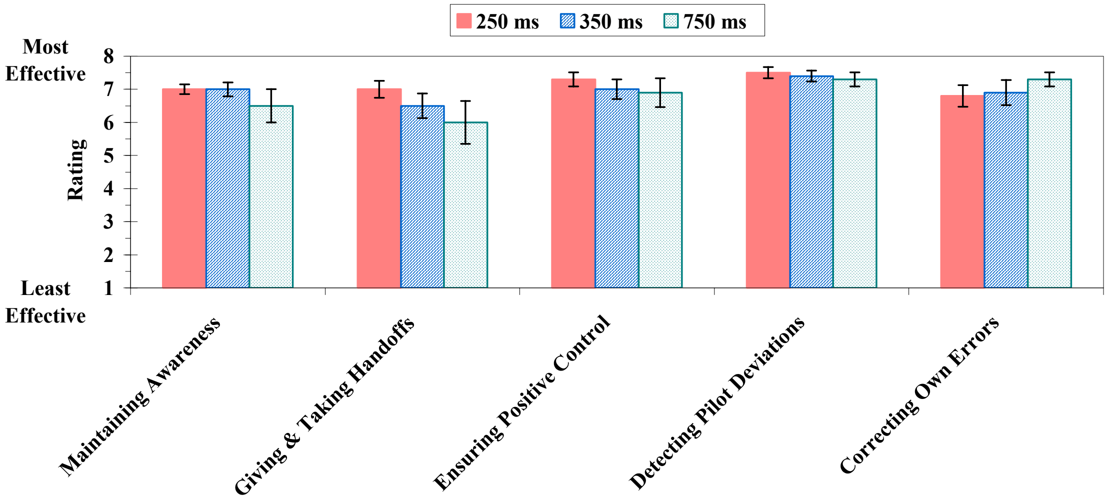
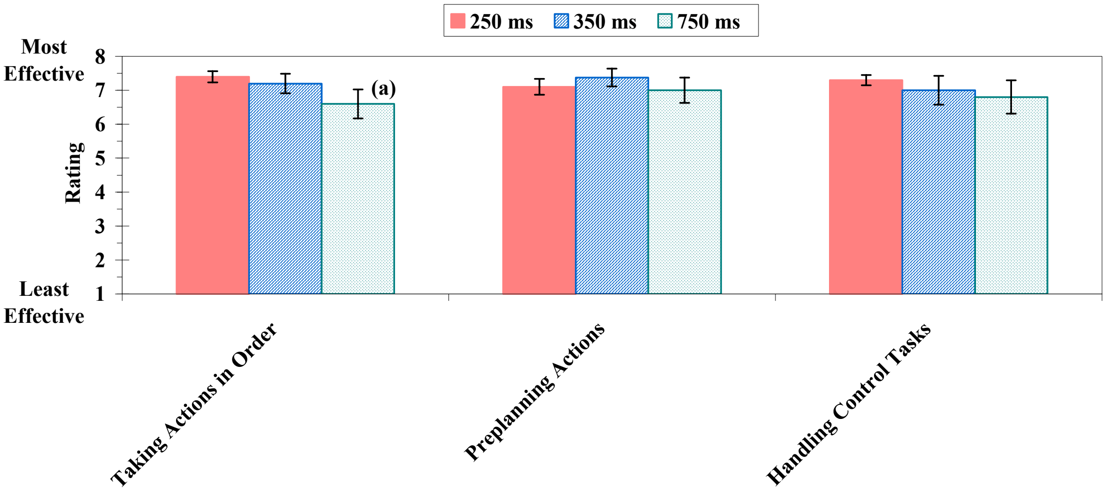


Figure 14. Mean observer ratings of controller performance for each Attention and Awareness subscale by delay condition.

Figure 15 shows observer mean ratings of controller performance on the subscales of the Prioritizing category for each of the delay conditions. The mean ratings were highly to very highly and ranged from 6.60 to 7.40 on an 8-point scale. Observer ratings on the Taking Actions in an Appropriate Order subscale were lower for the 750 ms delay compared to 250 ms delay, [*F*(2,18) = 5.32, *p* = .015]. On average, the observers rated the 750 ms delay as 0.80 points (i.e., 11%) lower than the 250 ms delay.



(a) 750 ms < 250 ms, (*p* < .05)

Figure 15. Mean observer ratings of controller performance for each Prioritizing subscale by delay condition.

Figure 16 shows observer mean ratings of controller performance on the subscales of the Providing Control Information category for each of the delay conditions. The mean ratings were highly to very highly effective and ranged from 6.63 to 7.50 on an 8-point scale. There were no significant differences between the delay conditions for any of the subscale ratings.

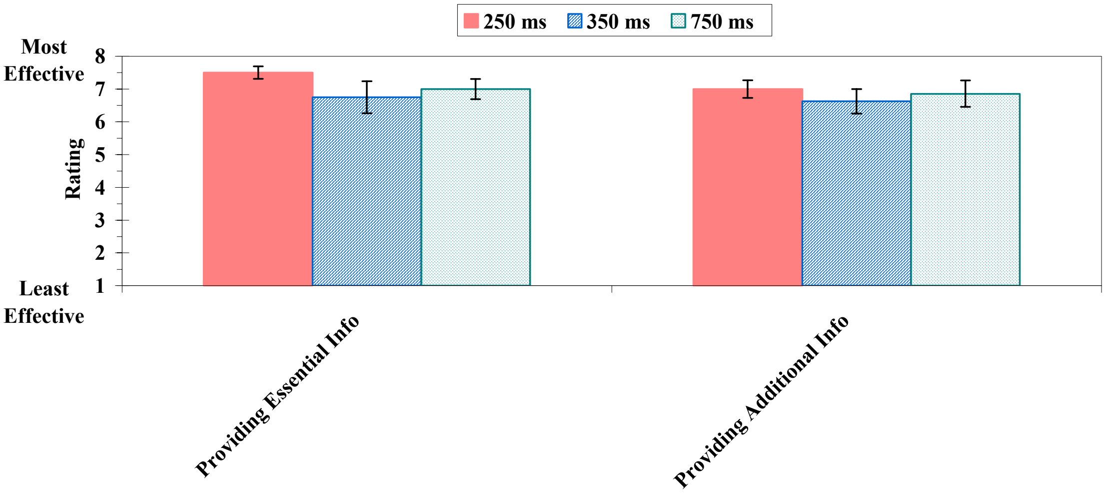


Figure 16. Mean observer ratings of controller performance for each Providing Control Information subscale by delay condition.

Figure 17 shows observer mean ratings of controller performance on the subscales of the Technical Knowledge category for each of the delay conditions. The mean ratings were very highly effective and ranged from 7.10 to 7.50 on an 8-point scale. There were no significant differences between the delay conditions for any of the subscale ratings.

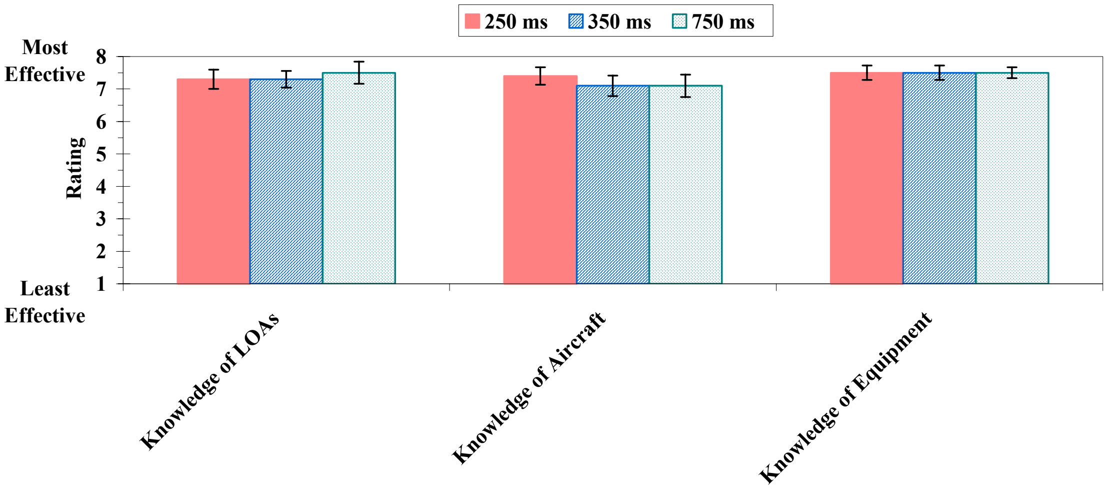


Figure 17. Mean observer ratings of controller performance for each Technical Knowledge subscale by delay condition.

Figure 18 shows observer mean ratings of controller performance on the subscales of the Communicating category for each of the delay conditions. The mean ratings were very highly effective and ranged from 7.00 to 7.50 on an 8-point scale. There were no significant differences between the delay conditions for any of the subscale ratings.

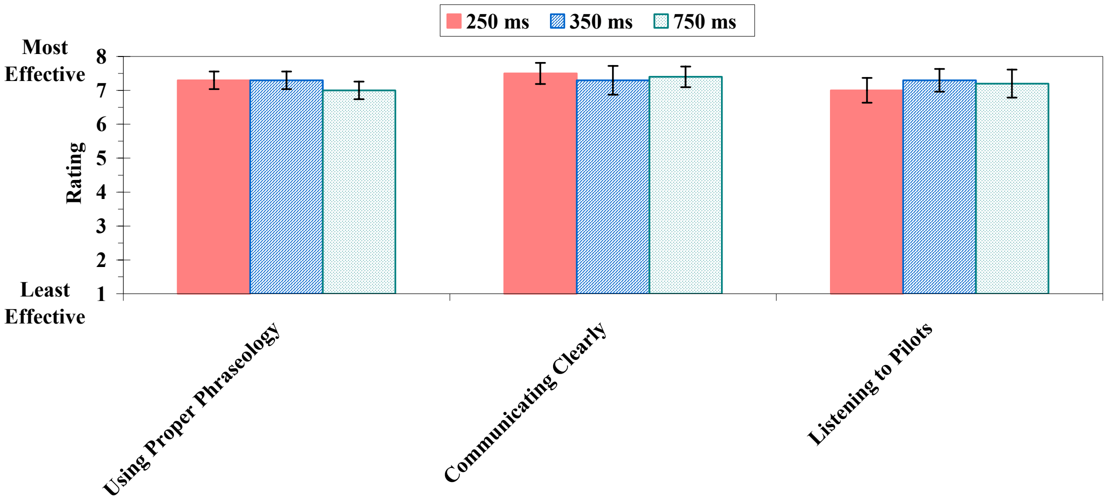


Figure 18. Mean observer ratings of controller performance for each Communicating subscale by delay condition.

Figure 19 shows observer mean ratings for the occurrence of key events for each of the delay conditions. The events were rare to normal in occurrence; mean ratings ranged from 1.30 to 2.50 on a 5-point scale. There were no significant differences between the delay conditions for any of the occurrence ratings.

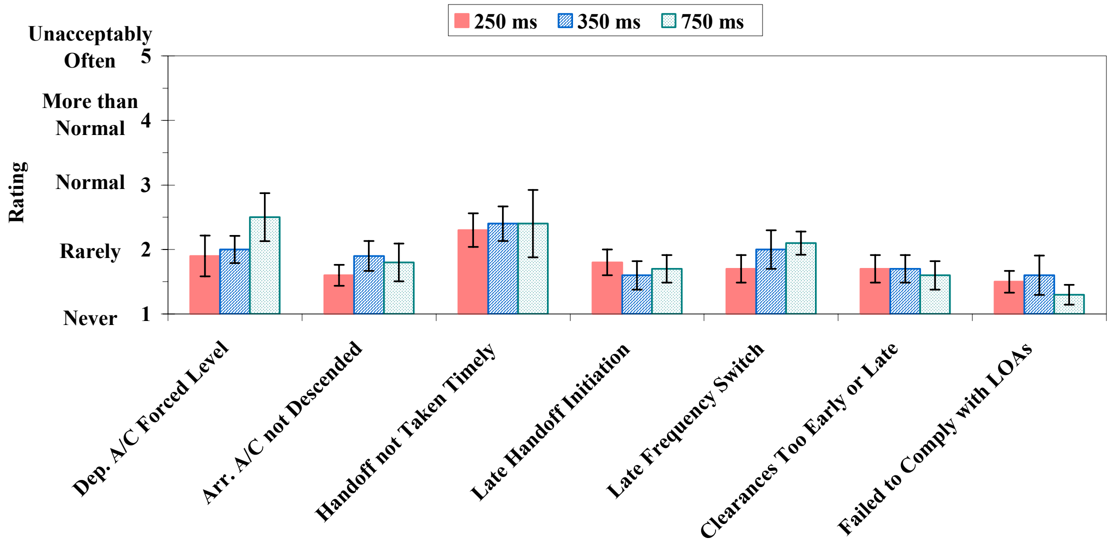


Figure 19. Mean observer ratings for the occurrence of key events by delay condition.

Figure 20 shows observer mean overall ATC safety ratings for each of the delay conditions. The observers rated the overall safety as acceptable; mean ratings ranged from 3.00 to 3.20 on a 4-point scale. There were no significant differences between the delay conditions for any of the safety ratings.

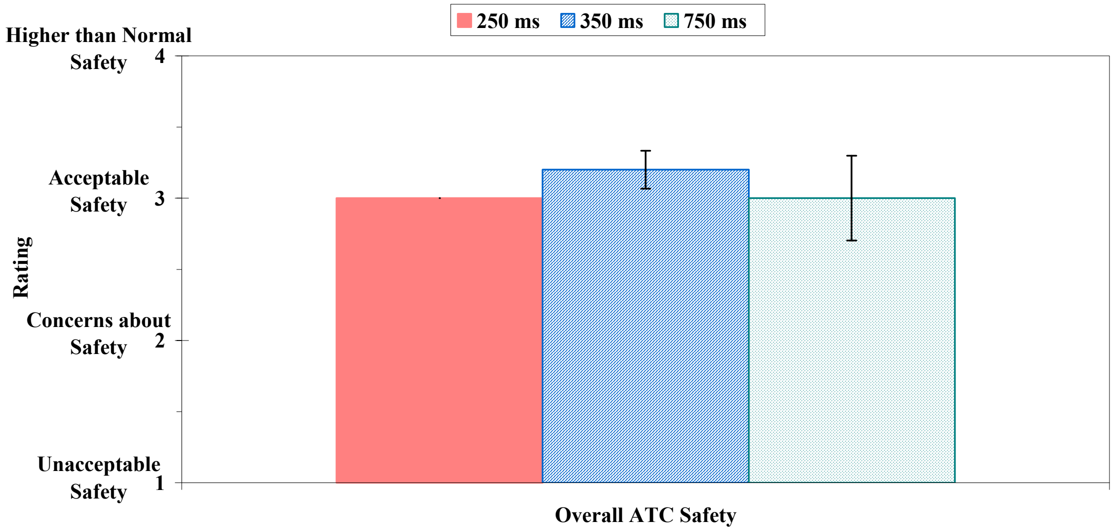


Figure 20. Mean observer overall ATC safety ratings by delay condition.

3.6 Controller Operational Acceptability

Figure 21 shows controller mean acceptability ratings and confidence levels for each of the delay conditions. The controllers rated the communications system as acceptable with moderate to negligible deficiencies; mean ratings ranged from 7.50 to 8.10 on a 10-point scale. There were no significant differences between the delay conditions for any of the acceptability ratings. Most of the confidence ratings were high (60%) and the rest were medium (40%).

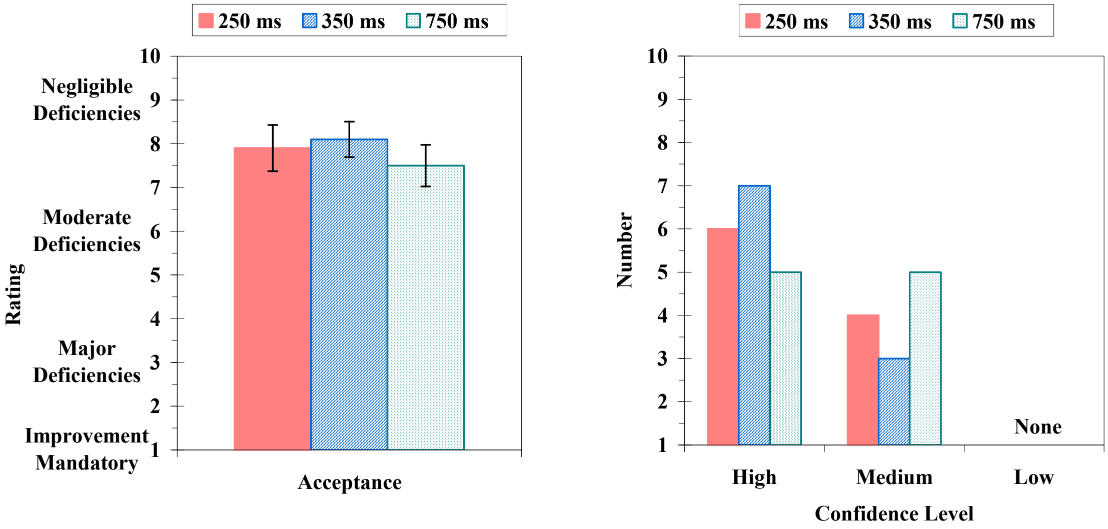


Figure 21. Mean acceptability ratings and confidence levels by delay condition.

3.7 Exit Questionnaire Ratings and Participant Comments

In the Exit Questionnaire, the controllers rated the controller override as very important for effective ATC performance (*M* = 8.7, *SD* = 1.42); importance ratings ranged from 6 to 10 (where 1 = None At All to 10 = A Great Deal). In the comments section of the Exit Questionnaire, one participant remarked that “being able to block the pilots and get transmissions through was an outstanding benefit.” Another participant commented that “[I] liked the override function.” In the final group discussion at the end of the simulation, controllers generally agreed that controller override was a good feature that had benefits for controllers. However, several controllers commented that they tried to avoid overriding pilots.

Also in the Exit Questionnaire, 9 of the 10 controllers indicated they noticed a difference in the communication delays. When asked to order the problems from longest to shortest delay, only 5 of 7 controllers who responded at all were able to correctly identify the problem with the longest delay. Only 2 of the 5 controllers who attempted to order all three delay conditions were able to correctly identify the problem with the shortest delay. In addition, 4 out of the 10 controllers indicated they adjusted their communications or control strategy because of the delays. One participant commented that “with each increase in delay, I had to change technique to compensate (i.e., starting aircraft down earlier, planning farther ahead for actions). [I] had to increase speech rate. [I] felt like I was behind more with the longer delays and it increased stress levels.” Another participant said, “I allowed a little [more] time before speaking [with the longest delay].”

Table 5 shows the controller ratings of simulation realism. The controllers rated the simulation realism as moderate; mean ratings ranged from 6.1 to 6.9 on a 10-point scale. Finally, the controllers rated the ATWIT workload rating technique as low in interference (*M* = 2.9, *SD* = 3.03); however, interference ratings ranged from 1 to 9 (where 1 = None At All, 10 = A Great Deal).

Table 5. Controller Ratings of Simulation Realism

Question	Extremely Unrealistic ①②③④⑤⑥⑦⑧⑨⑩ Extremely Realistic		
Rate the realism of the overall simulation experience compared to actual ATC operations.	Mean = 6.2	SD = 1.69	Range (3 – 8)
Rate the realism of the simulation DSR hardware compared to actual DSR equipment.	Mean = 6.2	SD = 2.10	Range (3 – 8)
Rate the realism of the simulation DSR software compared to actual DSR functionality.	Mean = 6.5	SD = 1.84	Range (4 – 9)
Rate the realism of the simulation traffic scenarios compared to actual NAS traffic.	Mean = 6.1	SD = 1.37	Range (3 – 7)
Rate the realism of the simulation generic airspace compared to actual NAS airspace.	Mean = 6.9	SD = 1.52	Range (4 – 9)

4. Discussion and Conclusions

To evaluate alternative VDL3 communication delays on controllers, we conducted a high fidelity controller-in-the-loop simulation of the en route ATC environment. We collected both objective and subjective data measuring various aspects of ATC performance that could be affected by communication delays. The results indicated that there were no differences between the 250 ms and the 350 ms delay condition in any of the measures we collected. There were no differences between 250 ms and 350 ms delays for aircraft flight times, distances, altitude, heading, and airspeed changes. In addition, there were no differences between 250 ms and 350 ms delays for controller communications, workload, situational awareness, and system acceptance. Finally, there were no differences between 250 ms and 350 ms delays for any of the SME observer ratings.

However, the 750 ms delay condition showed differences in the objective communications measures and controller ratings of the impact of communications. Controllers made fewer transmissions and occupied the communication channel for a shorter duration during the 750 ms delay compared to 250 ms and 350 ms delays. This result indicates that the controllers were using compensatory strategies for coping with the 750 ms delay. In the 750 ms delay, controllers may have made fewer optional ATC service transmissions (e.g., calling traffic) and may have talked faster or omitted words to shorten transmissions. Controller ratings of the impact of communications provide supporting evidence for these interpretations of the objective communications data. Controller ratings indicated that the 750 ms delay interfered with optional ATC services and speech clarity more than the 250 ms and 350 ms delays. Also, the controller ratings indicated that speech rate increased somewhat in the 750 ms delay condition, but this result was not statistically significant.

The objective communications data indicated that about 7% - 12% of pilot transmissions were overridden and that overrides occurred much more frequently in the 750 ms delay condition compared to the 250 ms and 350 ms delays. Most of the controller overrides of pilot transmissions were unintentional and occurred during the communications delay before the controller could hear the pilot speaking. In an analog radio system, the controller overrides would result in stepped-on transmissions that usually require retransmission from both the controller and pilot. However, the VDL3 controller override feature allows the controller transmission to be sent clearly, although the pilot may have to retransmit. In this manner, the controller override helped to mitigate the negative effects of communication delays for controllers. Controllers emphasized this point and the importance of controller override in their ratings at the end of the study and in the final group discussion.

This study was a high fidelity, controller-in-the-loop simulation of VDL3 system features including communication delays, controller override, antiblocking, and transmit status indicator. We used realistic controller workstations and field controllers as participants. The study was not as realistic from the pilot perspective. No line pilots participated, and we did not use realistic flight deck simulators. More important, we configured the simulation for three simulation pilots to support each controller position and maneuver from four to eight aircraft at a time. The limited number of pilots flying in each scenario may have constrained the level of channel contention observed in this study. For example, if every aircraft in the simulation was operated by a pilot, there could be an increased incidence of blocked or overridden pilot transmissions. Therefore, additional research is needed to examine the impact of communication delays, controller override, antiblocking, and transmit status indicator on operational communications from the pilot perspective.

The results of the study support the following conclusions and recommendations.

- The VDL3 communications system with controller override, antiblocking, and a transmit status indicator can be implemented with a 350 ms delay without causing problems for controllers.
    - There were no significant differences between the 250 ms and 350 ms conditions.
    - Controllers rated the override feature as very important for effective ATC performance.
- Additional research is needed to examine the impact of communication delays, controller override, antiblocking, and transmit status indicator on pilots.

Acronyms

ARTCC	Air Route Traffic Control Center
ATC	Air Traffic Control
ATCS	Air Traffic Control Specialist
ATS	Air Traffic Service
ATWIT	Air Traffic Workload Input Technique
CARS	Controller Acceptance Rating Scale
DESIREE	Distributed Environment for Simulation, Rapid Engineering, & Experimentation
DSR	Display System Replacement
ERP	Engineering Research Psychologist
FAA	Federal Aviation Administration
NAS	National Airspace system
NASA	National Aeronautics and Space Administration
NEXCOM	Next Generation Air-Ground Communications
RDHFL	Research, Development, and Human Factors Laboratory
SATCS	Supervisory Air Traffic Control Specialist
SME	Subject Matter Expert
TGF	Target Generator Facility
TLX	Taskload Index
VDL3	Very High Frequency Digital Link Mode 3
VHF	Very High Frequency
WJHTC	William J. Hughes Technical Center

References

Buckley, E. P., DeBaryshe, B. D., Hitchner, N., & Kohn, P. (1983). *Methods and measurements in real-time Air Traffic Control system simulation* (DOT/FAA/CT-83/26). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

Cardosi, K. M. (1993). Time required for transmission of critical Air Traffic Control messages in an en route environment. *The International Journal of Aviation Psychology*, *3*(4), 303-313.

Child, J., Cleve, R., & Grable, M. (1989). *Evaluation of low data rate voice CODECS for Air Traffic Control applications* (DOT/FAA/CT-TN89/13). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

Data Link Benefits Study Team. (1995). *User benefits of two-way data link ATC communications: Aircraft delay and flight efficiency in congested en route airspace* (DOT/FAA/CT-95/4). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

Dehel, T., Grable, M., & Child, J. (1989). *Phase II testing and evaluation of low data rate voice CODEC equipment* (DOT/FAA/CT-TN89/49). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

Federal Aviation Administration. (2000). *Requirements document for next generation air/ground communications system (NEXCOM): Segment 1*. Washington, DC: Author.

Federal Aviation Administration. (2001, July). *NEXCOM human factors plan*. Unpublished report. Washington, DC: Author.

Farncombe, D. (1997). *Investigations into the operational effects of the VDL Mode 3 voice throughput delay* (AMCP WG D/7 WP 53). Madrid, Spain: Aeronautical Mobile Communications Panel Working Group D.

Farncombe, D., & MacBride, J. J. (2000). *VDS Mode 3 Flight Tests using the AMBE-ATC-10B Vocoder* (AMCP WG D/7 – WP/UK4). Montreal, Canada: Aeronautical Mobile Communications Panel Working Group D.

Fujimori, T., & Ueno, M. (1999). *Results of Vocoder Evaluation in Japan* (AMCP WG D/10 WP 9). Hawaii, USA: Aeronautical Mobile Communications Panel Working Group D.

Guttman, J., & Stein, E. S. (1997). *En route generic airspace evaluation* (DOT/FAA/CT-TN97/7). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.

Guttman, J., Stein, E. S., & Gromelski, S. (1995). *The influence of generic airspace on Air Traffic Controller performance* (DOT/FAA/CT-TN95/38). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

Hart, S. G., & Staveland, L. E. (1987). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: North-Holland.

Kitawaki, N., & Itoh, K. (1991). Pure delay effects on speech quality in telecommunications. *IEEE Journal on Selected Areas in Communications*, *9*(4), 586-593.

LaDue, J., Sollenberger, R. L., Belanger, B., & Heinze, A. (1997). *Human factors evaluation of vocoders for Air Traffic Control environments – Phase I: Field evaluation* (DOT/FAA/CT-TN97/11). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.

Lee, K. K., Kerns, K., & Bone, R. (2001). Development and validation of the controller acceptance rating scale (CARS): Evidence from empirical research. *Proceedings of the 4th USA/Europe Air Traffic Management R&D Seminar*. Santa Fe, NM.

Nadler, E., Mengert, P., DiSario, R., Sussman, D. D., Grossberg, M., & Spanier, G. (1993). Effects of satellite- and voice-switching equipment transmission delays on Air Traffic Control communications. *The International Journal of Aviation Psychology*, *3*(4), 315-325.

Rantanen, E., McCarley, J., Xu, X., & Yeakel, S. (2001). *The effect of audio throughput delay of the NEXCOM system on Air Traffic Controller performance and workload* (Technical Report ARL-01-19/FAA-01-5). Savoy, IL: University of Illinois at Urbana-Champaign.

Renaud, P., Fistas, N., Brugere, C., & Garcia, A. (1999). *Results of vocoder intelligibility analysis for European languages* (AMCP WG D/11 – WP). Toulouse, France: Aeronautical Mobile Communications Panel Working Group D.

RTCA DO-225. (1994). *VHF air-ground communications system improvements alternatives study and selection of proposals for future action*. Washington, DC: RTCA, Inc.

Sollenberger, R. L., LaDue, J., Carver, B., & Heinze, A. (1997). *Human factors evaluation of vocoders for Air Traffic Control environments – Phase II: ATC simulation* (DOT/FAA/CT-TN97/25). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.

Sollenberger, R. L., Stein, E. S., & Gromelski, S. (1996). *The development and evaluation of a behaviorally based rating form for the assessment of Air Traffic Controller performance* (DOT/FAA/CT-TN96/16). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.

Stein, E. S. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

Stein, E. S. (1991). Evaluating Air Traffic Controller workload using real time person in the loop simulation. *Journal of Air Traffic Control, 33*(4), 55-58.

Stein, E. S., & Buckley, E. P. (1992). Unpublished Simulation Variables.

Suganuma, M. (1997). *Investigation of influences on ATC communications by the delay time generated in voice communications using VDL* (AMCP WG D7 WP 59). Madrid, Spain: Aeronautical Mobile Communications Panel Working Group D.

Vardaman, J. J., & Stein, E. S. (1998). *The development and evaluation of a behaviorally based rating form for the assessment of en route Air Traffic Controller performance* (DOT/FAA/CT-TN98/5). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.

Wickens, C. D. (1998). Common sense statistics. *Ergonomics in Design, 6*(4), 18-22.

Yuditsky, T., Sollenberger, R. L., Della Rocco, P. S., Friedman-Berg, F., & Manning, C. A. (2002). *Application of color to reduce air traffic complexity* (DOT/FAA/CT-TN02/17). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.

APPENDIX A

Informed Consent Form

I, _____________________________, understand that this study, entitled "The Effect of Voice Communications Latency In High Density, Communications-Intensive Airspace" is sponsored by the Federal Aviation Administration and is being directed by Dr. Randy Sollenberger.

Nature and Purpose:

I have been recruited to volunteer as a participant in this project. The purpose of the study is to determine the effects of alternative Very High Frequency Digital Link Mode 3 (VDL3) communications delays in a high-fidelity, controller-in-the-loop simulation. The results of the study will be used to establish VDL3 delay performance requirements.

Experimental Procedures:

Participants will arrive at the simulation laboratory in groups of three controllers for each 2-day simulation session. Each participant will work independent traffic scenarios that do not involve handoffs with other participants. The first day of the simulation will consist of a project briefing, airspace training, and three 60-minute practice scenarios. The second day of the simulation will consist of three 60-minute test scenarios and a final study debriefing. Participants will work from 8:00 AM to 4:30 PM each day with a rest break after each scenario and a lunch break.

Participants will control traffic under three different communications delay parameters. After each scenario, the controllers will complete questionnaires to evaluate the impact of the alternative communications delay parameters on controller workload and acceptance. In addition, subject-matter experts will make over-the-shoulder observations during the simulation to further assess the delay parameters. Finally, an automated data collection system will record system operations and generate a set of standard ATC simulation measures, which include safety, capacity, efficiency, and communications measures. The simulation will be audio-video recorded in case researchers need to re-examine any important simulation events.

Discomfort and Risks:

I understand that I will not be exposed to any foreseeable risks or intrusive measurement techniques.

Confidentiality:

My participation is strictly confidential, and no individual names or identities will be recorded or released in any reports.

Benefits:

I understand that the only benefit to me is that I will be able to provide the researchers with valuable feedback and insight into the effects of different communications delay parameters. My data will help the FAA to establish VDL3 delay performance requirements.

Participant Responsibilities:

I am aware that to participate in this study I must be a certified professional controller who is qualified at an air traffic control facility and holds a current medical certificate. I will control traffic and answer any questions asked during the study to the best of my abilities. I will not discuss the content of the experiment with anyone until the study is completed.

Participant's Assurances:

I understand that my participation in this study is completely voluntary and I have the freedom to withdraw at any time without penalty. I also understand that the researchers in this study may terminate my participation if they feel this to be in my best interest. I understand that if new findings develop during the course of this research that may relate to my decision to continue participation, I will be informed.

I have not given up any of my legal rights or released any individual or institution from liability for negligence.

Dr. Sollenberger has adequately answered all the questions I have asked about this study, my participation, and the procedures involved. I understand that Dr. Sollenberger or another member of the research team will be available to answer any questions concerning procedures throughout this study.

If I have questions about this study or need to report any adverse effects from the research procedures, I will contact Dr. Sollenberger at (609) 485-7169.

Compensation and Injury:

I agree to immediately report any injury or suspected adverse effect to Dr. Randy Sollenberger at (609) 485-7169. Local clinics and hospitals will provide any treatment, if necessary. I agree to provide, if requested, copies of all insurance and medical records arising from any such care for injuries/medical problems.

Signature Lines:

I have read this informed consent form. I understand its contents, and I freely consent to participate in this study under the conditions described. I understand that, if I want to, I may have a copy of this form.

Research Participant:____ Date:__________

Investigator:____ Date:__________

Witness:____ Date:__________

APPENDIX B

Background Questionnaire

Instructions:

This questionnaire is designed to obtain information about your background and experience as an air traffic control specialist (ATCS). The information will be used to describe the participants in this study as a group. You will not be identified by name.

Indicate your response by filling in the circle with a pen or pencil (or mark with an X), writing on the blank line, or circling the percentage number where appropriate. Some rating scales have labels at either end and 10 numbered circles that represent different levels of response. Fill in <u>one</u> circle with a pen or pencil (or mark with an X) to indicate your level of response.

Demographic Information and Experience

1. What is your **gender**?	○ Male	○ Female

2. Will you be **wearing corrective lenses** during this experiment?	○ Yes	○ No

3. What is your **age**?	_____ years _____ months

4. How long have you worked as an **ATCS (include both FAA and military experience)**?	_____ years _____ months

5. How long have you worked as an **ATCS for the FAA**?	_____ years _____ months

6. How long have you been a **Certified Professional Controller** (or Full Performance Level Controller)?	_____ years _____ months

7. How long have you **actively controlled traffic** in the en route environment?	_____ years _____ months

8. How many of the **past 12 months** have you actively controlled traffic?	_____ months

Sector Characteristics

9. Rate the **complexity** of the **sector** that you **work most frequently**.	Low Complexity	①②③④⑤⑥⑦⑧⑨⑩	High Complexity

10. Provide the approximate **dimensions** of your **most frequently worked sector**.	_____ nautical miles wide	_____ nautical miles long

11. Does your **most frequently worked sector** contain **special use** or **restricted airspace**?	○ Yes	○ No

12. Describe the **traffic type** of your **most frequently worked sector** by assigning a **percentage** to the following categories:	
(A). Transitional Traffic to/from a Major Airport	0 10 20 30 40 50 60 70 80 90 100
(B). High Altitude En Route (FL240 and above)	0 10 20 30 40 50 60 70 80 90 100
(C). Low Altitude En Route (FL230 and below)	0 10 20 30 40 50 60 70 80 90 100

13. Describe the **traffic mix** of your **most frequently worked sector** by assigning a **percentage** to the following categories (percentages must sum to 100%):	
(A). Air Carriers/Corporate Jets	0 10 20 30 40 50 60 70 80 90 100
(B). Air Taxi/Commuters	0 10 20 30 40 50 60 70 80 90 100
(C). Cargo	0 10 20 30 40 50 60 70 80 90 100
(D). General Aviation	0 10 20 30 40 50 60 70 80 90 100
(E). Military	0 10 20 30 40 50 60 70 80 90 100
(F). Other, specifiy ____________________	0 10 20 30 40 50 60 70 80 90 100
Sum Total	100%

14. Rate how often you use **paper flight strips**.	Never	①②③④⑤⑥⑦⑧⑨⑩	Always

General Ratings

15. Rate your current **skill as an ATCS**.	Not Skilled	①②③④⑤⑥⑦⑧⑨⑩	Extremely Skilled
16. Rate your current **level of stress**.	Not Stressed	①②③④⑤⑥⑦⑧⑨⑩	Extremely Stressed
17. Rate your **level of motivation** to participate in this study.	Not Motivated	①②③④⑤⑥⑦⑧⑨⑩	Extremely Motivated

APPENDIX C

Post-Scenario Questionnaire

Instructions:

Please answer the following questions based upon your experience in the scenario just completed. The rating scales have labels at either end and 10 numbered circles that represent different levels of response. Fill in one circle with a pen or pencil (or mark with an X) to indicate your level of response.

Overall Performance, Workload, Situational Awareness, and Simulation Ratings

1. Rate your **overall level of ATC performance** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
2. Rate your **overall level of situational awareness** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
3. Rate your **situational awareness for current aircraft locations** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
4. Rate your **situational awareness for projected aircraft locations** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
5. Rate your **situational awareness for potential aircraft loss-of-separation** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
6. Rate your **workload due to communications** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High
7. Rate the **performance of the simulation pilots** in terms of their responding to your control instructions and providing readbacks.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
8. Rate the **difficulty** of this scenario.	Extremely Easy	①②③④⑤⑥⑦⑧⑨⑩	Extremely Difficult

Specific Communications Delay Ratings

9. To what extent did the **delays interfere with the effectiveness of your communications** during this scenario?	None At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

10. To what extent did the **delays interfere with your control strategy (or style)** during this scenario?	None At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

11. To what extent did the **delays interfere with the timing of your control instructions** during this scenario?	None At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

12. To what extent did the **delays interfere with your providing optional (or courtesy) air traffic services** during this scenario?	None At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

13. To what extent did the **delays interfere with your using correct phraseology** during this scenario?	None At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

14. To what extent did the **delays interfere with your speech quality (or clarity)** during this scenario?	None At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

15. To what extent did the **delays increase your speech rate** during this scenario?	None At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

16. Do you have any additional comments or clarifications about your experience in the simulation?

____

____

____

____

____

____

____

NASA-TLX Ratings

Definitions

Mental Demand – how much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Were your tasks easy or demanding, simple or complex, exacting or forgiving?

Physical Demand – how much physical activity was required (e.g., data entry, strip marking, talking, pointing, etc.)? Were your tasks easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Temporal Demand – how much time pressure did you feel due to the rate or pace at which your tasks occurred? Was the pace slow and leisurely or rapid and frantic?

Performance – how successful do you think you were in accomplishing the goals of your tasks? How satisfied were you with your performance in accomplishing these goals?

Effort – how hard did you have to work (mentally and physically) to accomplish this level of performance?

Frustration – how insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel in performing your tasks?

17. Rate your **mental demand** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

18. Rate your **physical demand** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

19. Rate your **temporal demand** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

20. Rate your **performance** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

21. Rate your **effort** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

22. Rate your **frustration** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

23. Do you have any comments or clarifications about these NASA-TLX questions?

____

____

APPENDIX D

Controller Acceptance Rating Scale

Guidelines For Numerical Rating

Procedure:

1. Start at the top left-hand corner of the page.
2. Answer each yes/no question according to the problem that you just controlled.
3. Use the definitions below to make the judgments.
4. Circle **one** number from 1 to 10 that **best** reflects your experience in the run you just controlled.
5. Provide a Confidence Rating according to the definitions below.
6. Please add comments to explain your rating.

Definitions

System:

The system is taken to mean **everything** being rated:

- The controller's performance,
- The performance of the ground automation system, including the performance of Host computer, the performance of the air-ground communications system, and
- Pilot performance

When evaluating the overall system and considering pilot performance:

1. If pilot response is exceptionally bad (e.g., not very responsive) over several aircraft, then this could lead to a poorer picture of how well the overall system could perform. This should be reflected in the **confidence rating**. But to the extent that the controller procedures were affected by bad pilot response, that should be considered in the **numerical rating**.
2. If pilot response is bad, but the controller procedures seem to react especially poorly or especially well in adapting to the pilot situation, then this should be considered in the **numerical rating**.

Confidence:

The **Confidence Rating** should describe confidence in the numerical acceptability rating itself. **It is <u>not</u> a rating of how confident one is <u>about</u> the air-ground communications system**.

The **Confidence Rating does** answer the question, **"How confident am I that the rating I just made is an accurate one, reflecting the overall system performance, based on the amount of information I had available to me?"**

The **Confidence Rating should reflect the amount of information you think you had available to you in making your overall rating**. It should also reflect problems that you encountered that are not necessarily an indication of how the air-ground communications system performed. As in the example above, a pilot that is especially unresponsive and uncooperative which results in a difficult traffic situation could mean that any problems encountered in the traffic situation could be due to more than just air-ground communication system performance, the pilot response is also a factor. How much a factor is reflected in the confidence rating.

There are 3 Levels of Confidence Rating:

High Confidence

You were able to account for the traffic events that occurred. You are very certain what problems or benefits could be due to air-ground communications, the traffic situation, etc., and can therefore provide a rating that really reflects how well the air-ground communications system performed.

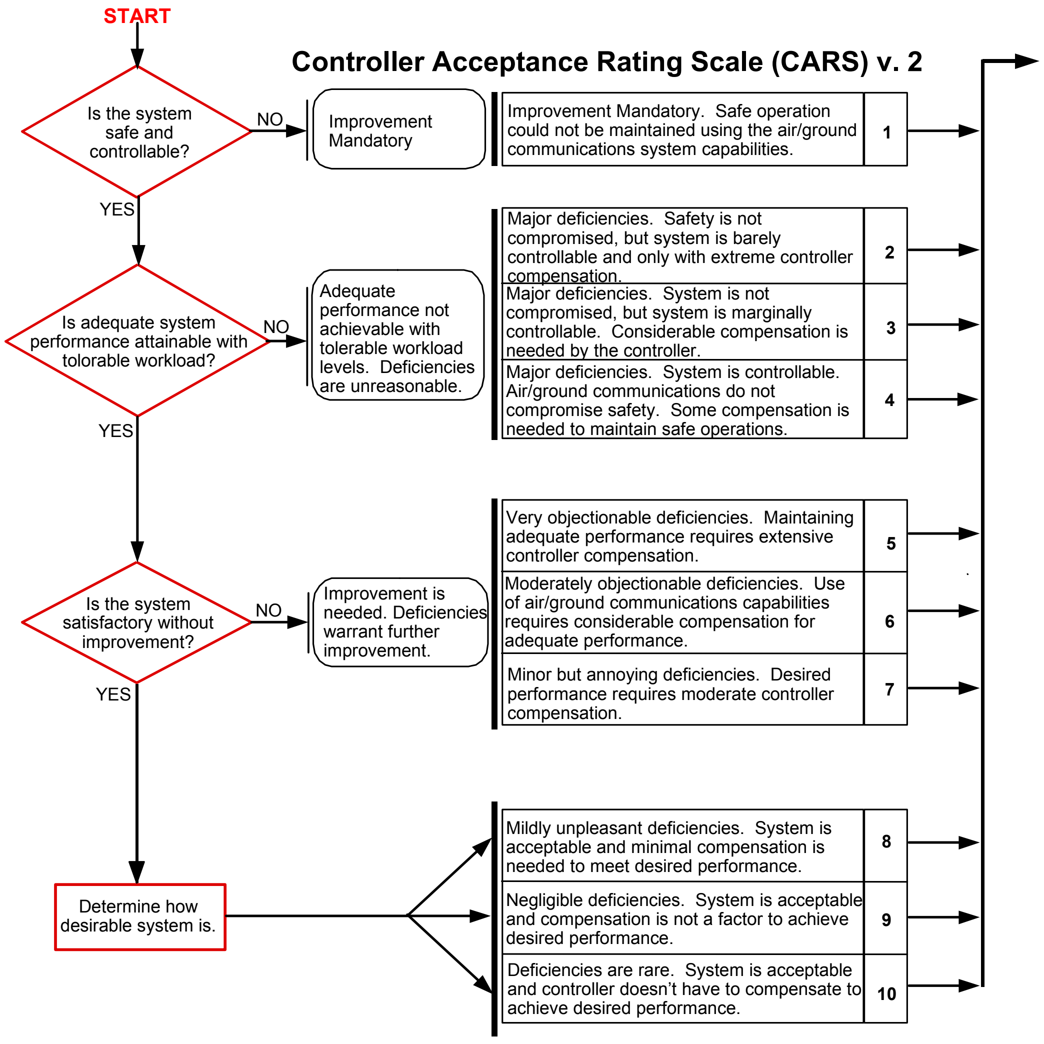
Moderate Confidence

You were able to account for some of the traffic outcome. You are somewhat certain what problems or benefits could be due to communications, the traffic situation, etc. There is some uncertainty about how well the air-ground communications system performed, given the overall situation. You have some reservations about the accuracy of your numerical rating.

Low Confidence

It was difficult to account for the traffic outcome. There is a great deal of uncertainty about the performance of the air-ground communications system, and how you were able to work within the whole system. You have many reservations about the accuracy of your numerical rating because of external factors that you cannot adequately account for.

Controller Acceptance Rating Scale (CARS) v. 2



Improvement Mandatory. Safe operation could not be maintained using the air/ground communications system capabilities.	**1**
Major deficiencies. Safety is not compromised, but system is barely controllable and only with extreme controller compensation.	**2**
Major deficiencies. System is not compromised, but system is marginally controllable. Considerable compensation is needed by the controller.	**3**
Major deficiencies. System is controllable. Air/ground communications do not compromise safety. Some compensation is needed to maintain safe operations.	**4**
Very objectionable deficiencies. Maintaining adequate performance requires extensive controller compensation.	**5**
Moderately objectionable deficiencies. Use of air/ground communications capabilities requires considerable compensation for adequate performance.	**6**
Minor but annoying deficiencies. Desired performance requires moderate controller compensation.	**7**
Mildly unpleasant deficiencies. System is acceptable and minimal compensation is needed to meet desired performance.	**8**
Negligible deficiencies. System is acceptable and compensation is not a factor to achieve desired performance.	**9**
Deficiencies are rare. System is acceptable and controller doesn't have to compensate to achieve desired performance.	**10**

Provide Confidence Rating:

A B C

Comments:

____
____
____
____
____
____
____
____
____
____
____
____
____
____
____

Circle appropriate information below:

Team: 1 2 3 4 5 6

Run: 1 2 3 4 5 6

Controller Postion: R D

APPENDIX E

Subject Matter Expert Observer Rating Form

Observer Code _________ Date _________
Controller _________ Scenario _________

INSTRUCTIONS

This form is designed to be used by supervisory air traffic control specialists to evaluate the effectiveness of controllers working in simulation environments. SATCSs will observe and rate the performance of controllers in several different performance dimensions using the scale below as a general purpose guide. Use the entire scale range as much as possible. You will see a wide range of controller performance. Take extensive notes on what you see. Do not depend on your memory. Write down your observations. Space is provided after each scale for comments. You may make preliminary ratings during the course of the scenario. However, wait until the scenario is finished before making your final ratings and remain flexible until the end when you have had an opportunity to see all the available behavior. At all times please focus on what you actually see and hear. This includes what the controller does and what you might reasonably infer from the actions of the pilots. Try to avoid inferring what you think may be happening. If you do not observe relevant behavior or the results of that behavior, then you may leave a specific rating blank. Also, please write down any comments that may help improve this evaluation form. Do not write your name on the form itself. You will not be identified by name. An observer code known only to yourself and the researchers conducting this study will be assigned to you. The observations you make do not need to be restricted to the performance areas covered in this form and may include other areas that you think are important.

ASSUMPTIONS

ATC is a complex activity that contains both observable and unobservable behavior. There are so many complex behaviors involved that no observational rating form can cover everything. A sample of the behaviors is the best that can be achieved, and a good form focuses on those behaviors that controllers themselves have identified as the most relevant in terms of their overall performance. Most controller performance is at or above the minimum standards regarding safety and efficiency. The goal of the rating system is to differentiate performance above this minimum. The lowest rating should be assigned for meeting minimum standards and also for anything below the minimum since this should be a rare event. It is important for the observer/rater to feel comfortable using the entire scale and to understand that all ratings should be based on behavior that is actually observed.

Rating Scale Descriptors

Remove this Page and keep it available while doing ratings

SCALE	QUALITY	SUPPLEMENTARY
1	**Least Effective**	Unconfident, Indecisive, Inefficient, Disorganized, Behind the power curve, Rough, Leaves some tasks incomplete, Makes mistakes
2	**Poor**	May issue conflicting instructions, Doesn't plan completely
3	**Fair**	Distracted between tasks
4	**Low Satisfactory**	Postpones routine actions
5	**High Satisfactory**	Knows the job fairly well
6	**Good**	Works steadily, Solves most problems
7	**Very Good**	Knows the job thoroughly, Plans well
8	**Most Effective**	Confident, Decisive, Efficient, Organized, Ahead of the power curve, Smooth, Completes all necessary tasks, Makes no mistakes

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW

II - MAINTAINING ATTENTION AND SITUATION AWARENESS

III – PRIORITIZING

IV – PROVIDING CONTROL INFORMATION

V – TECHNICAL KNOWLEDGE

VI – COMMUNICATING

1. Departing a/c forced to level
2. Arriving a/c not descended
3. H/O not taken (timely manner)
4. Late H/O (late initiation)
5. Late switch (freq)
6. Controller checks for pilot on frequency (missed call in)
7. Retransmissions (2x) and “say again” (SA)
8. Aircraft left on heading
9. Unexpected conflict alerts (CA) and separation errors (SEP)
10. Controller’s reaction to pilot requests (comment on each of the 3 requests)

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW

1.	Maintaining Separation and Resolving Potential Conflicts • using control instructions that maintain appropriate aircraft and airspace separation • detecting and resolving impending conflicts early • recognizing the need for speed restrictions and wake turbulence separation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
2.	Sequencing Aircraft Efficiently • using efficient and orderly spacing techniques for arrival, departure, and en route aircraft • maintaining safe arrival and departure intervals that minimize delays	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
3.	Using Control Instructions Effectively/Efficiently • providing accurate navigational assistance to pilots • issuing economical clearances that result in need for few additional instructions to handle aircraft completely • ensuring clearances require minimum necessary flight path changes	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
4.	Overall Safe and Efficient Traffic Flow Scale Rating	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

II - MAINTAINING ATTENTION AND SITUATION AWARENESS

5.	Maintaining Awareness of Aircraft Positions • avoiding fixation on one area of the radar scope when other areas need attention • using scanning patterns that monitor all aircraft on the radar scope	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
6.	Giving and Taking Handoffs in a Timely Manner • ensuring that handoffs are initiated in a timely manner • ensuring that handoffs are accepted in a timely manner • ensuring that handoffs are made according to procedures	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
7.	Ensuring Positive Control • tailoring control actions to situation • using effective procedures for handling heavy, emergency, and unusual traffic situations	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
8.	Detecting Pilot Deviations from Control Instructions • ensuring that pilots follow assigned clearances correctly • correcting pilot deviations in a timely manner	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
9.	Correcting Own Errors in a Timely Manner • acting quickly to correct errors • changing an issued clearance when necessary to expedite traffic flow	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
10.	Overall Attention and Situation Awareness Scale Rating	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

III – PRIORITIZING

11. Taking Actions in an Appropriate Order of Importance • resolving situations that need immediate attention before handling low priority tasks • issuing control instructions in a prioritized, structured, and timely manner	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
12. Preplanning Control Actions • scanning adjacent sectors to plan for future and conflicting traffic • studying pending flight strips in bay	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
13. Handling Control Tasks for Several Aircraft • shifting control tasks between several aircraft when necessary • communicating in timely fashion while sharing time with other actions	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
14. Marking Flight Strips while Performing Other Tasks • marking flight strips accurately while talking or performing other tasks • keeping flight strips current	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
15. Overall Prioritizing Scale Rating	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

IV – PROVIDING CONTROL INFORMATION

16. Providing Essential Air Traffic Control Information • providing mandatory services and advisories to pilots in a timely manner • exchanging essential information	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
17. Providing Additional Air Traffic Control Information • providing additional services when workload permits • exchanging additional information	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
18. Providing Coordination • providing effective and timely coordination • using proper point-out procedures	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
19. Overall Providing Control Information Scale Rating	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

V – TECHNICAL KNOWLEDGE	
20. Showing Knowledge of LOAs and SOPs • controlling traffic as depicted in current LOAs and SOPs • performing handoff procedures correctly	**1 2 3 4 5 6 7 8**
21a. Showing Knowledge of Aircraft Capabilities and Limitations • using appropriate speed, vectoring, and/or altitude assignments to separate aircraft with varied flight capabilities • issuing clearances that are within aircraft performance parameters	**1 2 3 4 5 6 7 8**
21b. Showing Effective Use of Equipment • updating data blocks • using equipment capabilities	**1 2 3 4 5 6 7 8**
22. Overall Technical Knowledge Scale Rating	**1 2 3 4 5 6 7 8**

VI – COMMUNICATING	
23. Using Proper Phraseology • using words and phrases specified in the 7110.65 • using phraseology that is appropriate for the situation • using minimum necessary verbiage	**1 2 3 4 5 6 7 8**
24. Communicating Clearly and Efficiently • speaking at the proper volume and rate for pilots to understand • speaking fluently while scanning or performing other tasks • ensuring clearance delivery is complete, correct and timely • speaking with confident, authoritative tone of voice	**1 2 3 4 5 6 7 8**
25. Listening to Pilot Readbacks and Requests • correcting pilot readback errors • acknowledging pilot or other controller requests promptly • processing requests correctly in a timely manner	**1 2 3 4 5 6 7 8**
26. Overall Communicating Scale Rating	**1 2 3 4 5 6 7 8**

	Frequency of Occurrence				
Task	**Never Occurred**	Rarely Occurred	**Occurred, but within Normal Limits of Operational Acceptability**	Occurred More Than Normal	**Occurred Unacceptably Often**
1. Departing aircraft forced to level	①	②	③	④	⑤
2. Arriving aircraft not descended	①	②	③	④	⑤
3. Handoff not taken (in timely manner)	①	②	③	④	⑤
4. Late handoff (initiation)	①	②	③	④	⑤
5. Late switch (on frequency)	①	②	③	④	⑤
6. Issued clearances earlier or later than appropriate	①	②	③	④	⑤
7. Failed to comply with Letters of Agreement	①	②	③	④	⑤

	Overall ATC Performance			
Task	Operations were Unsafe and Unacceptable	**Operational Safety was not Compromised, but I had Safety Concerns**	Operations were Typical of this Type of Sector with Acceptable Safety	**The Margin of Safety was Higher Than Normal for this Type of Sector**
8. Overall Operational Assessment of ATC Performance	①	②	③	④

If you marked ① or ② for your overall operational assessment of ATC performance, please explain your rating below. Thoroughly describe the incidents or factors that influenced your judgment.

APPENDIX F

Exit Questionnaire

Instructions:

Please answer the following questions based upon your overall experience in the simulation. The rating scales have labels at either end and 10 numbered circles that represent different levels of response. Fill in one circle with a pen or pencil (or mark with an X) to indicate your level of response.

Simulation Realism and Research Apparatus Ratings

1. Rate the **realism of the overall simulation experience** compared to actual ATC operations.	Extremely Unrealistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic
2. Rate the **realism of the simulation DSR hardware** compared to actual DSR equipment.	Extremely Unrealistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic
3. Rate the **realism of the simulation DSR software** compared to actual DSR functionality.	Extremely Unrealistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic
4. Rate the **realism of the simulation traffic scenarios** compared to actual NAS traffic.	Extremely Unrealistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic
5. Rate the **realism of the simulation generic airspace** compared to actual NAS airspace.	Extremely Unrealistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic
6. To what extent did the **ATWIT online workload rating technique interfere** with your ATC performance?	None At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

7. Do you have any comments or suggestions for improvement about our simulation capability?

____

____

____

____

Final Communications Questions

8. How important was the **controller override for effective ATC performance**?	Not At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

9. Did you **notice any differences in the length of the communications delays**?	○ Yes	○ No

If yes, please explain the differences between the first problem (P1), second problem (P2), and third problem (P3) (e.g., which was the longest delay, shortest delay, and middle delay).

10. Did you **adjust your communications or control strategies because of the delays**?	○ Yes	○ No

If yes, please explain the differences between the first problem (P1), second problem (P2), and third problem (P3).

11. Is there anything about the study that we should have asked or that you would like to comment about?