te technical note techn

Human Factors Evaluation of Vocoders for Air Traffic Control (ATC) Environments Phase II: ATC Simulation

Randy L. Sollenberger, Ph.D., ACT-530
James La Due, Ph.D., SRC
Brian Carver, ATCS
Annmarie Heinze, SRC

December 1997

DOT/FAA/CT-TN97/25

Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

U.S. Department of Transportation
Federal Aviation Administration

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

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/CT-TN97/25		

4. Title and Subtitle Human Factors Evaluation of Vocoders for Air Traffic Control Environments Phase II: ATC Simulation	**5. Report Date** December 1997 **6. Performing Organization Code** ACT-530
7. Author(s) Randy Sollenberger, Ph.D., ACT-530, James La Due, Ph.D., SRC, Brian Carver, ATCS, and Annmarie Heinze, SRC	**8. Performing Organization Report No.** DOT/FAA/CT-TN97/25
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** DTFA03-94-C-00042
12. Sponsoring Agency Name and Address Federal Aviation Administration Communications & Infrastructure Branch William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	**13. Type of Report and Period Covered** Technical Note **14. Sponsoring Agency Code** ACT-330

15. Supplementary Notes

16. Abstract

Vocoders offer a potential solution to radio congestion by digitizing human speech and compressing the signal to achieve low bandwidth voice transmissions. A reduction in bandwidth will allow the addition of more communication channels to the system and reduce radio congestion. This air traffic control simulation study is the second phase of a research effort to compare the effectiveness of two 4.8 kbps vocoders (designated as A and B for test purposes) with the current analog radio communication system.
Sixteen air traffic controllers from Level 5 Terminal Radar Approach Controls participated in the study and performed 12 one-hour traffic scenarios over 3 days of testing. Scenarios consisted of medium and high traffic volumes designed to produce different levels of controller taskload. The communications configuration allowed each simulation pilot to transmit with jet, propeller, or helicopter background noises. The results indicated that the vocoders did not affect controller workload or performance. In general, intelligibility and acceptability ratings were highest for analog radio, slightly lower for vocoder B, and lowest for vocoder A. In addition, intelligibility and acceptability ratings were highest for jet background noise, slightly lower for propeller background noise, and lowest for helicopter background noise. Controller taskload had no effect on intelligibility and acceptability. This human factors evaluation indicated that both vocoders were highly intelligible and acceptable for air traffic control environments. Even the least preferred vocoder did not substantially interfere with controller performance. This study suggests that vocoder technology could replace the current analog radio system in the future.

17. Key Words	18. Distribution Statement
Air Traffic Control, Communications, Vocoders, Simulation, Performance, Human Factors	This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	48	

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

Acknowledgement

The authors would like to gratefully acknowledge the engineering support and cooperation of ACT-330 team members, Rodney Guishard, James Eck, Edward Coleman, and John Petro. In addition, the authors would like to recognize the assistance of George Rowand (SRC) and William Belanger (EPA).

Table of Contents

Page

Acknowledgement ... iii

Executive Summary ... vii

1. Introduction ... 1
   - 1.1 Background ... 1
   - 1.2 Purpose ... 1
   - 1.3 Scope ... 1

2. Method ... 2
   - 2.1 Participants ... 2
   - 2.2 Simulation ... 2
   - 2.3 Airspace ... 4
   - 2.4 Traffic Scenarios ... 6
   - 2.5 Design ... 6
     - 2.5.1 Independent Variables ... 6
     - 2.5.2 Dependent Variables ... 7
   - 2.6 Training ... 7
   - 2.7 Procedure ... 7

3. Results ... 11
   - 3.1 System Effectiveness Measures ... 12
   - 3.2 Observer Ratings ... 14
   - 3.3 Controller Ratings ... 16
   - 3.4 Final Questionnaire ... 23

4. Discussion and Conclusions ... 23

References ... 26

Appendixes

A - ATC System Effectiveness Measures
B - Observer Rating Form
C - Post-Scenario Questionnaire

List of Illustrations

Figures Page

1. Simulation Setup and Organization of Controllers, Observers, and Simulation Pilots 3
2. Communications and Aircraft Background Noise Considerations 5
3. Mean Number of Push-to-Talk Transmissions as a Function of Communication Equipment and Controller Taskload 13
4. Mean Duration of Push-To-Talk Transmissions as a Function of Communication Equipment and Controller Taskload..14
5. Mean Observer Rating for Listening to Pilot Readbacks and Requests as a Function of Communication Equipment and Controller Taskload...16
6. Taxonomy of Observer Comments as a function of the Communication Equipment...............17
7. Mean Air Traffic Workload Input Technique Ratings as a Function of Communication Equipment and Controller Taskload 19
8. Mean Intelligibility Ratings for all Transmissions as a Function of Communication Equipment and Controller Taskload..20
9. Mean Acceptability Ratings for all Transmissions as a Function of Communication Equipment and Controller Taskload..20

Tables Page

1. Representative ATC System Effectiveness Measures..8
2. Observation Form Rating Scales 8
3. Controllers' Subjective Ratings...9
4. Scenario Counterbalancing 10
5. Scenario Presentation Order 11
6. *F* Statistics Obtained from the Two-way ANOVA Performed on the System Effectiveness Measures 13
7. *F* Statistics Obtained from the Two-way ANOVA Performed on the Observer Ratings 15
8. *F* Statistics Obtained from the Two-way ANOVA Performed on the Controller Ratings 18
9. Degrees of Freedom, Mean Squares, and *F* Statistics Obtained from the Three-way ANOVA Performed on the Intelligibility Ratings Ratings 21
10. Mean Intelligibility Ratings, *F* Statistics Obtained from the Analysis of Simple Main Effects, and Tukey HSD Post Hoc Comparisons..21
11. Degrees of Freedom, Mean Squares, and *F* Statistics Obtained from the Three-way ANOVA Performed on the Acceptability Ratings...22
12. Mean Acceptability Ratings, *F* Statistics Obtained from the Analysis of Simple Main Effects, and Tukey HSD Post Hoc Comparisons..22
13. Exit Questionnaire Ratings..23

Executive Summary

Researchers at the William J. Hughes Technical Center Research Development and Human Factors Laboratory conducted a human factors evaluation of current vocoder technology with controllers in a real-time air traffic control (ATC) simulation. In the phase I study, the researchers presented auditory recordings to controllers who provided intelligibility and acceptability ratings as well as objective understandability responses. The purpose of phase II was to confirm the findings of the previous study and investigate a larger number of performance measures under realistic ATC conditions. The study compared the effectiveness of two vocoders (denoted as vocoder A and vocoder B for test purposes) relative to the current analog radio communication system. The researchers examined the effects of controller taskload and aircraft background noises on each communication system.

Sixteen air traffic controllers from Level 5 Terminal Radar Approach Controls (TRACONs) participated in the study. The controllers arrived at the laboratory in pairs, and the researchers conducted two independent simulations simultaneously. The experimental apparatus consisted of a high-fidelity ATC simulator with a voice communication link between each controller and a team of trained simulation pilots. Each controller operated a radar position without assistance. Each of the simulation pilots transmitted with a different aircraft background noise and responded to controller clearances appropriate to the aircraft type. The background noises included jet aircraft, propeller aircraft, and helicopters.

The controllers performed 12 one-hour traffic scenarios over 3 days of testing. Scenarios consisted of medium and high traffic volumes designed to produce different levels of controller taskload. Medium taskload scenarios consisted of 48 aircraft, and high taskload scenarios consisted of 60 aircraft appearing within a 1-hour period. Over the course of the experiment, each participant used all three communication systems and worked a different set of four traffic scenarios with each system. The researchers selected a generic Level 5 TRACON sector for phase II that was developed and validated in previous research.

The experimental design included several different ATC performance measurements. The laboratory automated data collection system produced a large set of system effectiveness measures that provided objective indicators of safety, capacity, and efficiency. An air traffic control specialist (ATCS) made over-the-shoulder ratings using an observation form specifically designed for ATC performance evaluation research. Controllers provided overall intelligibility and acceptability ratings for each communication system and individual ratings under each type of aircraft background noise. In addition, the controllers provided ratings of their mental, physical, and temporal workload after each scenario using the National Aeronautical and Space Administration Taskload Index procedure. The system also collected real-time workload ratings from controllers every 5 minutes using the Air Traffic Workload Input Technique. The researchers did not inform the participants which communication system was operating during each scenario.

The results indicated that the vocoders did not affect controller workload or system safety, capacity, and efficiency. As in the first phase of the study, subjective intelligibility ratings were slightly higher than acceptability ratings. However, unlike phase I, the intelligibility and

acceptability ratings in phase II showed a high degree of correlation. In general, overall intelligibility and acceptability ratings were highest for analog radio, only slightly lower for vocoder B, and lowest for vocoder A. The results indicated an interaction between the communication equipment and aircraft background noises for both intelligibility and acceptability ratings. For jet and propeller background noises, intelligibility and acceptability were the lowest for vocoder A, but there were no significant differences between analog radio and vocoder B. For helicopter background noise, intelligibility and acceptability were the highest for analog radio, but there were no significant differences between vocoder A and vocoder B.

Controller taskload did not affect intelligibility and acceptability ratings but had very strong effects on the other dependent measures. Safety, capacity, and efficiency indicators showed that controllers committed more separation errors, completed more flights, and issued more clearances in high taskload scenarios. Observer and controller performance ratings were generally lower in high taskload scenarios. Mental, physical, temporal, and overall workload were higher in high taskload scenarios.

The intelligibility and acceptability results of the simulation agreed with the findings of the phase I study. Both phases suggest that vocoder B is very comparable to analog radio and vocoder A is less intelligible and acceptable to controllers. Although the researchers collected a large number of objective ATC performance measures and other subjective ratings, there were no other differences between the three communication systems. These results suggest that even the least preferred vocoder did not have substantial detrimental effects on controller performance. However, both phases of the study have examined a limited set of factors that could potentially influence the effectiveness of vocoders. Future research should investigate additional issues.

1. Introduction

1.1 Background

Radio congestion is a major problem facing the air traffic control (ATC) system today. The Federal Aviation Administration (FAA) currently maintains 25 kHz bandwidth between analog radio channels in the ATC system. A reduction in this bandwidth will allow the addition of more channels to the system and reduce radio congestion. Vocoders (voice coders) offer one possible solution for reducing channel bandwidth. A successful implementation of vocoders, however, requires that the speech produced by them be intelligible and acceptable for air traffic controllers and pilots. This study investigates vocoder human factors issues using a real-time ATC simulation to evaluate the effectiveness of vocoders under realistic ATC conditions.

Vocoders are a digital communication technology that converts human speech into a compressed digital format that radios can transmit. The compression process depends upon a speech model to produce signals that sound like the original speech. The result is that vocoders can transfer speech signals at very low bit rates over a digital communication link.

Vocoders offer advantages over the current analog radio communication system. The proposed bit rate of 4.8 kbps can potentially increase the number of available ATC communication channels by a factor of four. In addition, digital technologies offer improved security for communications and solutions to the problems of stuck microphones and "stepped on" transmissions. Vocoders do have limitations, however. Because of approximations made in the compression process, vocoder transmissions may sound somewhat different from what controllers have come to expect.

1.2 Purpose

The purpose of this phase of the vocoder study was to conduct a human factors evaluation of current vocoder technology with air traffic controllers in a real-time ATC simulation. The researchers intended the simulation to confirm the intelligibility and acceptability findings of the first phase (La Due, Sollenberger, Belanger, & Heinze, 1997) and to investigate a larger number of performance measures under realistic ATC conditions. As in the first phase, the present study compared the effectiveness of two vocoders (denoted as vocoder A and vocoder B for test purposes) relative to the current analog radio communication system. In addition, the researchers investigated the effects of controller taskload and aircraft background noises on each communication system.

1.3 Scope

The researchers limited the study to controller reception of pilot transmissions. Pilot reception of controller transmissions is a separate issue that would require certified pilots and other resources that were beyond the scope of this study but may be examined in a future study. As in the first phase of this study, the researchers set the bit error rate of the vocoders at 10^{-3}, which has been the standard in most vocoder research (Child, Cleve, & Grable, 1989; Dehel, Grable, & Child, 1989). The bit error rate determines the frequency of bit errors produced in the transmissions and represents another source of signal degradation other than the compression process in vocoder communications. The researchers also set the volume level of the aircraft background noises at 90 dB, which is typical for the cockpits of most civil aviation jet, propeller, and helicopter aircraft. The results of this study may not be applicable to military aircraft that have louder cockpits. The present study did not systematically investigate the sex of the speakers as in the first phase. However, the researchers did record the sex of the simulation pilots and controllers participating in the study.

2. Method

2.1 Participants

Sixteen male air traffic controllers from 13 Level 5 Terminal Radar Approach Controls (TRACONs) volunteered for this study. All participants were full performance level (FPL) controllers, and all but one had actively controlled traffic for the past 12 months. Each controller completed an initial questionnaire to describe the background characteristics of participants in the study. Controllers ranged in age from 32 to 52 years old (Mean = 38.94, *SD* = 4.88), and ranged in experience from 8 to 34 years of active service (Mean = 17.06, *SD* = 6.69). Additionally, controllers provided self-ratings of three personal attributes that could affect simulation performance. The rating scale ranged from 1 (meaning low/poor) to 10 (meaning high/good) on each question. The attributes included enthusiasm to participate (Mean = 8.81, *SD* = 1.17), health (Mean = 8.56, *SD* = 1.46), and prior knowledge of vocoders (Mean = 2.50, *SD* = 1.79).

2.2 Simulation

Researchers conducted the simulation in the Research Development and Human Factors Laboratory (RDHFL) at the FAA William J. Hughes Technical Center. The simulation equipment consisted of state-of-the-art controller workstations with large high-resolution displays, a voice communication system, networked computer resources, and ATCoach simulation software (copyright UFA Inc., 1992). Two human factors specialists and one current Level 5 TRACON air traffic control specialist (ATCS) conducted the simulation and observed the participants in the control room. A voice communication link to another room allowed controllers to issue ATC commands to a team of trained simulation pilots. The simulation pilots moved the aircraft radar targets using simple keyboard commands and communicated with the controllers using proper ATC phraseology.

The researchers printed and time-ordered flight progress strips in a strip bay before the start of each scenario. During the simulation, audio-visual equipment recorded the controllers' radar display, voice communications, and actions for future reference. The researchers conducted two independent simulations simultaneously. Each controller operated a radar position without assistance.

Figure 1 illustrates the overall setup and organization of the simulation pilots, controllers, and observer. In each of the independent sessions, one simulation pilot (denoted as A1 or B1) operated all aircraft using simple keyboard commands and did not communicate with controllers

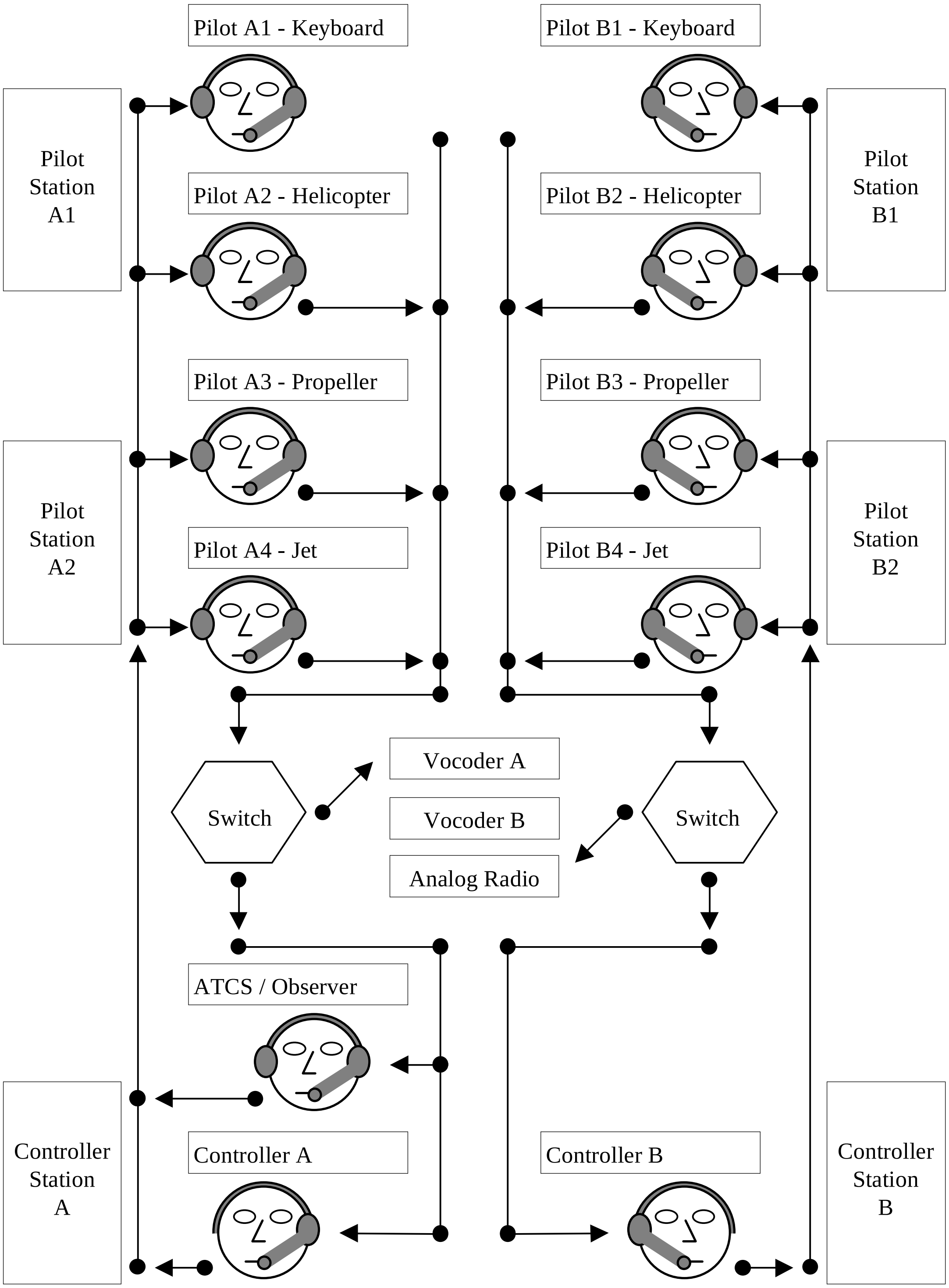

Figure 1. Simulation setup and organization of controllers, observers, and simulation pilots.

(denoted as A or B). Three other pilots communicated with the controllers. Each of these pilots transmitted with a different aircraft background noise and responded to controller clearances of the appropriate aircraft type. One pilot (denoted as A2 or B2) transmitted with a helicopter background noise, a second pilot (denoted as A3 or B3) transmitted with a propeller aircraft background noise, and the third pilot (denoted as A4 or B4) transmitted with a jet aircraft background noise. In addition to readbacks, the simulation pilots provided initial contact communications and replied to traffic advisories. The ATCS observed over the shoulder of one controller at a time for each scenario but switched to watching the other controller on alternate scenarios.

The researchers modified the laboratory communication system to incorporate the vocoders and a noise generator that produced realistic static in analog radio transmissions. The signal-to-noise ratio for analog radio transmissions was comparable to that produced at 50% of the service distance for ATC radio antennas. As illustrated in Figure 2, simulation pilots wore enclosed headsets, and when they keyed their microphones, the system produced aircraft background noise and side-tone in their headsets. The researchers adjusted the side-tone level so that the natural speaking volume of each pilot produced a voice signal that controllers heard above the background noise. The researchers set the volume level of all aircraft background noises at 90 dB. Pilot transmissions passed through one of the two vocoders or the analog radio simulator. The controllers heard aircraft background noises in all communications with pilots. Controllers wore open-ear headsets, and when they keyed their microphones, the system produced side-tone only in their headsets. The controllers' transmissions to the simulation pilots were always through a clear communication channel because pilot reception was not the focus of this study. The researchers recorded ATC background noise from Philadelphia TRACON and played the tape over the control room speakers while the controllers worked traffic.

2.3 Airspace

The research team selected a generic Level 5 TRACON sector that was developed and validated in a previous human factors simulation study (Guttman, Stein, & Gromelski, 1995). Generic airspace has several advantages relative to modeling an actual sector in simulations. The generic airspace was designed to provide a realistic Level 5 TRACON environment for controlling traffic and to be easy for controllers to learn. The generic sector consisted of easily remembered fix names and simplified operating procedures. Using generic airspace, researchers can select a cross-section of controllers from different air traffic facilities and quickly train them to operate in the airspace. Actual airspace is much more difficult for controllers from other facilities to learn. Using actual airspace, only a restricted sample of qualified controllers from a single facility can participate in a simulation. Additionally, it can typically take months of training for controllers to become qualified in an actual sector that is unfamiliar.

GENERA (GEN), the generic TRACON sector, was designed in a four-corner post configuration typical of most Level 5 TRACONs. Arrival aircraft entered the sector from the northwest, northeast, south, and southeast. Departure aircraft exited the sector to the north, east, west, and southwest. The sector consisted of a central major airport with parallel runways and three minor airports. In the actual simulation, only the right parallel runway was active, and the minor airports were not operational.

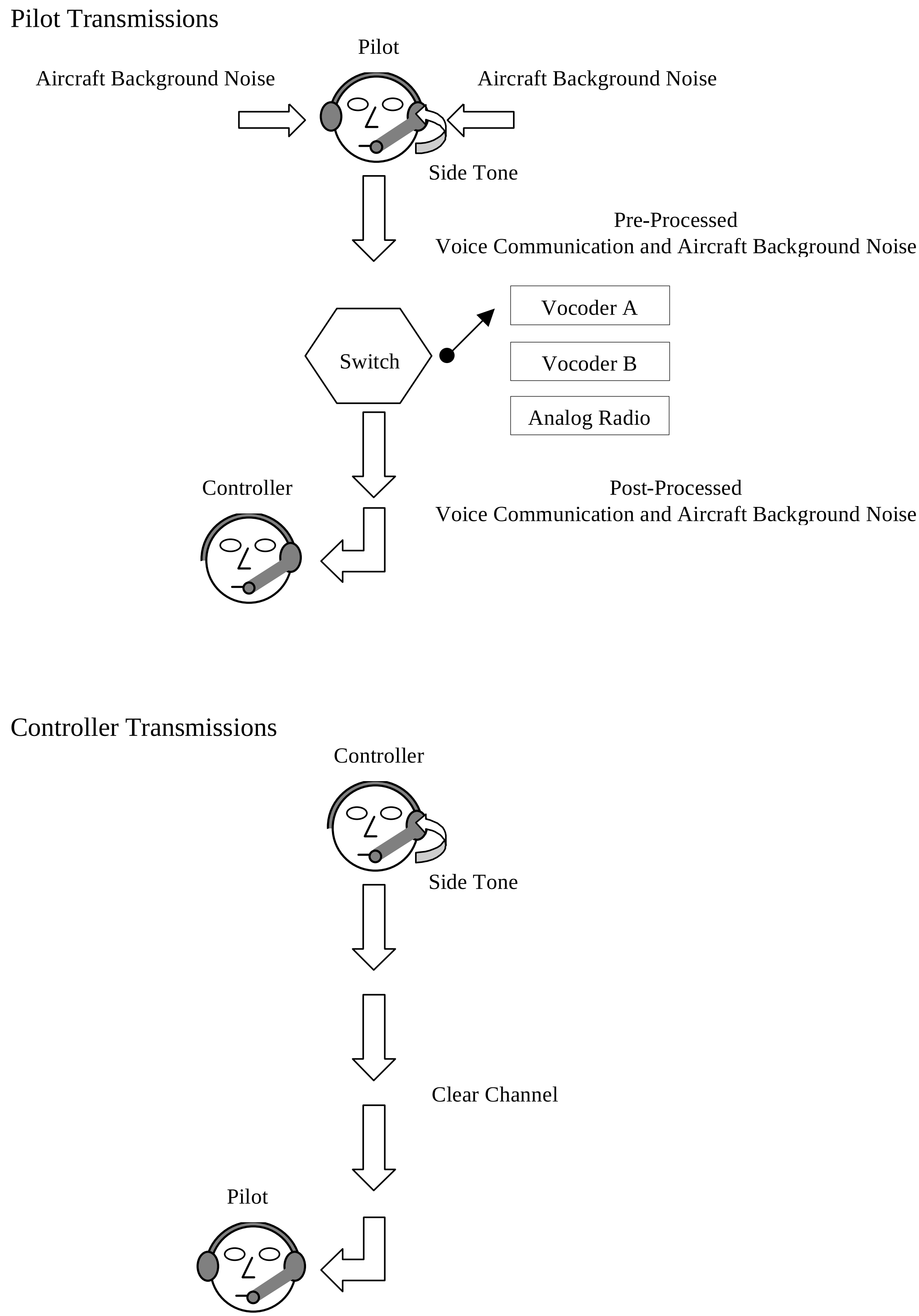

Figure 2. Communications and aircraft background noise considerations.

2.4 Traffic Scenarios

The human factors specialists and an ATCS constructed 12 air traffic scenarios for the simulation. Each scenario was 1 hour in duration and consisted of a mix of jet, propeller, and helicopter aircraft operating in Instrument Flight Rules (IFR) conditions. All scenarios started without any aircraft on the radar display. Then, aircraft steadily appeared, creating a buildup of traffic that maintained until the conclusion of the scenario. Designing scenarios with either a medium or high volume of traffic produced different levels of taskload. Medium taskload scenarios consisted of 48 aircraft appearing within a 1-hour period -- 34 arrivals and 14 departures. High taskload scenarios consisted of 60 aircraft appearing within a 1-hour period -- 42 arrivals and 18 departures. Three ATCSs pre-evaluated these aircraft numbers to ensure that they represented realistic traffic volumes for Level 5 facilities. The researchers designed the scenarios with different traffic flow characteristics to ensure that each scenario presented different ATC challenges for the controllers.

2.5 Design

2.5.1 Independent Variables

The main independent variable used in the simulation was the type of communication equipment. Each participant controlled different traffic scenarios using either vocoder A, vocoder B, or the analog radio simulator. The analog radio simulator was the “control” condition of the experiment that served as the standard of comparison for the vocoders. The second independent variable was the level of controller taskload that the researchers varied by designing scenarios with either a medium or high volume of traffic.

A third independent variable examined was the type of aircraft background noise. However, the researchers could not systematically manipulate aircraft background noise as other independent variables in the simulation. Although different aircraft background noises were included in pilot transmissions, the experimental design could not determine the individual effects of jet, propeller, and helicopter noises for most of the dependent measures. However, the researchers were able to examine controller’s subjective ratings of intelligibility and acceptability for the different aircraft background noises.

The experimental design can be summarized as a 3 x 2 within-subjects (or repeated measures) design with the factors of Equipment (vocoder A, vocoder B, analog radio) and Taskload (medium, high). For the intelligibility and acceptability ratings, the researchers conducted a 3 x 2 x 3 within-subjects analysis with the addition of Background Noise (jet, propeller, helicopter).

2.5.2 Dependent Variables

The RDHFL automated data collection system produces a large set of system effectiveness measures for ATC simulation research (Buckley, DeBaryshe, Hitchner, & Kohn, 1983; Stein & Buckley, 1992). Although researchers examined the entire set of measures, this study will report the results from a much smaller subset. Table 1 shows the subset of measures selected as representative indicators in the critical performance areas of safety, capacity, and efficiency (Appendix A lists the complete set of system effectiveness measures).

In addition to these objective performance measures, an ATCS observed controllers and made over-the-shoulder ratings of performance. The ATCS used an observation form specially designed for ATC performance evaluation research (Sollenberger, Stein, & Gromelski, 1997). Table 2 shows the 24 different rating scales of the observation form organized into 6 major performance categories (Appendix B displays the actual Observer Rating Form).

Finally, controllers provided intelligibility and acceptability ratings for the vocoders and analog radio simulator after each scenario. In addition, controllers provided self-ratings indicating their overall performance, situational awareness, and workload. Included in the ratings were workload scales based upon the National Aeronautical and Space Administration Taskload Index (NASA-TLX), a multi-dimensional workload assessment method (Hart & Staveland, 1988). During each scenario, controller workload was sampled using the Air Traffic Workload Input Technique (ATWIT), a real-time workload assessment method. Table 3 shows the ratings collected from controllers (Appendix C displays the actual Post-Scenario Questionnaire).

2.6 Training

Controllers participated in a training program to help them learn the generic airspace and become familiar with the simulation setup and procedures. The researchers developed a training manual that described the generic sector standard operating procedures (SOPs), letters of agreement (LOAs), sector layouts, arrival and departure routes, transfer of control points, and runway approach procedures. An ATCS reviewed the main points of the manual with controllers then illustrated the procedures while conducting special demonstration scenarios. In the remaining training time, controllers worked two 30-minute practice scenarios. The researchers did not intend the practice scenarios to be part of the communication equipment evaluation. Therefore, participants did not use the vocoders during practice and communicated using the analog radio simulator.

2.7 Procedure

The controllers arrived at the RDHFL in pairs for a week of simulation testing and evaluation. Monday and Friday were travel days. Tuesday, Wednesday, and Thursday consisted of project briefing, sector training, and simulation test scenarios. The participants worked from 8:00 AM to 4:30 PM with a 1-hour lunch period and three 10-minute breaks each day. The controllers completed a background questionnaire on the first day and a final questionnaire on the last day of the study. After each scenario, controllers completed a post-scenario questionnaire (see Appendix C).

Table 1. Representative ATC System Effectiveness Measures

I – SAFETY	
NSTCNF	Number of standard terminal conflicts
NLCNF	Number of ILS conflicts
II - CAPACITY	
NCOMP	Number of flights completed
NHAND	Number of flights handled
CMAV	Cumulative average of system activity/aircraft density
III - EFFICIENCY	
NPTT	Number of controller push-to-talk transmissions
DPTT	Duration of controller push-to-talk transmissions
NALT	Number of altitude clearances
NHDG	Number of heading clearances
NSPD	Number of airspeed clearances
DHAND	Duration of flights handled
DIST	Distance flown for flights

Table 2. Observation Form Rating Scales

I – MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW
1. Maintaining Separation and Resolving Potential Conflicts
2. Sequencing Arrival and Departure Aircraft Efficiently
3. Using Control Instructions Efficiently/Effectively
4. Overall Safe and Efficient Traffic Flow Scale Rating
II - MAINTAINING ATTENTION AND SITUATION AWARENESS
5. Maintaining Awareness of Aircraft Positions
6. Ensuring Positive Control
7. Detecting Pilot Deviations from Control Instructions
8. Correcting Own Errors in a Timely Manner
9. Overall Attention and Situation Awareness Scale Rating
III - PRIORITIZING
10. Taking Actions in an Appropriate Order of Importance
11. Preplanning Control Actions
12. Handling Control Tasks for Several Aircraft
13. Marking Flight Strips while Performing Other Tasks
14. Overall Prioritizing Scale Rating
IV – PROVIDING CONTROL INFORMATION
15. Providing Essential Air Traffic Control Information
16. Providing Additional Air Traffic Control Information
17. Overall Providing Control Information Scale Rating
V – TECHNICAL KNOWLEDGE
18. Showing Knowledge of LOAs and SOPs
19. Showing Knowledge of Aircraft Capabilities and Limitations
20. Overall Technical Knowledge Scale Rating
VI – COMMUNICATING
21. Using Proper Phraseology
22. Communicating Clearly and Efficiently
23. Listening to Pilot Readbacks and Requests
24. Overall Communicating Scale Rating

Table 3. Controllers' Subjective Ratings

1. Controller performance
2. Controller workload
3. Controller situation awareness
4. Simulation pilot performance
5. NASA-TLX, mental demand
6. NASA-TLX, physical demand
7. NASA-TLX, temporal demand
8. NASA-TLX, performance
9. NASA-TLX, effort
10. NASA-TLX, frustration
11a. Intelligibility, overall transmissions
11b. Acceptability, overall transmissions
12a. Intelligibility, jet transmissions
12b. Acceptability, jet transmissions
13a. Intelligibility, propeller transmissions
13b. Acceptability, propeller transmissions
14a. Intelligibility, helicopter transmissions
14b. Acceptability, helicopter transmissions
ATWIT, Air Traffic Workload Input Technique

Table 4 shows the scenario counterbalancing features of the experiment. The researchers assigned controllers to one of three groups (denoted A, B, or C). Each group of controllers used each of the three communication systems and worked a different set of four traffic scenarios with each system. Each set of scenarios consisted of two medium (e.g., M1 and M2) and two high (e.g., H1 and H2) taskload scenarios. An important feature of the experimental design to emphasize is that each controller worked each scenario only once. If controllers repeated the scenarios using different communication systems, the scenarios would have been easier to perform the second time due to familiarity with the traffic problems. Additionally, a different group of controllers worked each set of scenarios using different communication systems. This technique ensured that, if there were any especially easy or difficult scenarios, controllers worked them with each of the communication systems.

Table 5 shows the presentation order of the scenarios. The researchers randomly ordered the presentation of scenarios except for a few constraints. The two controllers in each pair (e.g., 1 and 2) used different communication systems at the same time because only one vocoder A, vocoder B, and analog radio simulator was available for the simulation. In addition, the two controllers worked different scenarios at the same time to avoid confusion from hearing each other issue clearances to the same aircraft. As indicated in the table, the ATCS alternated between the two controllers and observed only scenarios M1, M3, M5, H1, H3, and H5. The controllers did not work any of these scenarios simultaneously at the two positions.

Table 4. Scenario Counterbalancing

Group A Participant	Vocoder A				Vocoder B				Analog Radio			
1	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
2	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
3	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
4	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
5	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
6	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
Group B Participant	Vocoder B				Analog Radio				Vocoder A			
7	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
8	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
9	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
10	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
11	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
12	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
Group C Participant	Analog Radio				Vocoder A				Vocoder B			
13	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
14	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
15	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6
16	M1	M2	H1	H2	M3	M4	H3	H4	M5	M6	H5	H6

Note.
M1, M2, M3, M4, M5, and M6 are similar moderate traffic scenarios
H1, H2, H3, H4, H5, and H6 are similar high traffic scenarios

Table 5. Scenario Presentation Order

Participant	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
1	M5:R#	H4:B	M3:B#	M6:R	H5:R#	H2:A	M1:A#	M2:A	H3:B#	M4:B	H1:A#	H6:R
2	H2:A	H1:A#	M6:R	H3:B#	M2:A	M5:R#	M4:B	H5:R#	H6:R	M1:A#	H4:B	M3:B#
3	H5:R#	M2:A	H1:A#	H2:A	M5:R#	H6:R	H3:B#	H4:B	M3:B#	M6:R	M1:A#	M4:B
4	H2:A	M5:R#	H4:B	H3:B#	M4:B	M1:A#	M2:A	H1:A#	H6:R	M3:B#	M6:R	H5:R#
5	M5:R#	H2:A	M3:B#	M6:R	M1:A#	M4:B	H3:B#	H4:B	H5:R#	M2:A	H1:A#	H6:R
6	H4:B	H5:R#	M2:A	H1:A#	H6:R	M1:A#	M6:R	M5:R#	H2:A	M3:B#	M4:B	H3:B#
7	H1:B#	H6:A	M3:R#	M6:A	M5:A#	H4:R	M1:B#	H2:B	H5:A#	M2:B	H3:R#	M4:R
8	M6:A	M1:B#	M2:B	H3:R#	H4:R	H5:A#	H6:A	M3:R#	M4:R	M5:A#	H2:B	H1:B#
9	M5:A#	H6:A	M1:B#	H4:R	H5:A#	M4:R	M3:R#	M2:B	H3:R#	M6:A	H1:B#	H2:B
10	H2:B	H3:R#	M6:A	H1:B#	M4:R	M5:A#	M2:B	H5:A#	H6:A	M1:B#	H4:R	M3:R#
11	H5:A#	M6:A	H3:R#	M2:B	M1:B#	H6:A	M5:A#	M4:R	M3:R#	H2:B	H1:B#	H4:R
12	H2:B	H1:B#	H6:A	H3:R#	M4:R	M3:R#	M2:B	M1:B#	M6:A	M5:A#	H4:R	H5:A#
13	M3:A#	H6:B	H5:B#	M4:A	M1:R#	M2:R	M5:B#	H4:A	H3:A#	H2:R	H1:R#	M6:B
14	H6:B	H3:A#	H4:A	H1:R#	M6:B	H5:B#	H2:R	M5:B#	M2:R	M3:A#	M4:A	M1:R#
15	M5:B#	H2:R	M1:R#	H6:B	H3:A#	H4:A	H1:R#	M6:B	H5:B#	M2:R	M3:A#	M4:A
16	M4:A	M5:B#	H6:B	H1:R#	M6:B	H5:B#	H4:A	M3:A#	M2:R	H3:A#	H2:R	M1:R#

Note.
M1, M2, M3, M4, M5, and M6 are similar moderate traffic scenarios
H1, H2, H3, H4, H5, and H6 are similar high traffic scenarios
A, B, and R denote vocoder A, vocoder B, and analog radio, respectively
# indicates the ATCS observed the scenario

The researchers used ATWIT to assess controller workload as the participants conducted traffic. ATWIT provides an unobtrusive and reliable means for collecting controllers' workload ratings (Stein, 1985; Stein, 1991). A touch screen presented a workload rating scale and collected controllers' responses. Controllers indicated their current workload level by pressing one of the touch screen buttons labeled from 1 (indicating low workload) to 10 (indicating high workload). The system requested the controllers' input every 5 minutes by emitting several beeps and presenting the rating scale. Participants had 20 seconds to respond by pressing one of the 10 buttons. If controllers were too busy to respond within the allowed time, the system recorded a workload rating of 10 by default.

3. Results

The researchers used Analysis of Variance (ANOVA) to determine the effects of the communication equipment, controller taskload, and when possible, background noise on the dependent measures collected in the simulation. ANOVA is a statistical procedure for determining whether the differences between means are due to the independent (or treatment) variables or due to chance alone. The results of the analysis produce an *F* statistic and an associated *p* value. The *p* value is the probability that the differences in the means are due to chance alone. Researchers compare the *p* value to a selected significance level to determine if the treatment is statistically reliable or significant. A treatment with a *p* value greater than .05 is not statistically significant.

Researchers refer to the analyses associated with each independent variable as main effects and the analyses associated with combinations of variables as interaction effects. An interaction occurs when the effects of one variable are different depending upon the level of another variable. If an interaction is significant, the experimental design must be broken down into its basic components, referred to as simple main effects. One simple main effect involves the differences between the three communication systems for low taskload scenarios, and another involves the differences between the systems for high taskload scenarios. Researchers compute an *F* statistic for each simple main effect. Significant main effects or simple main effects with more than two treatment levels (e.g., vocoder A, vocoder B, and analog radio) must be analyzed by a post hoc comparison procedure to determine which levels are statistically different. In the present study, researchers used the Tukey Honestly Significant Difference (HSD) test for all post hoc comparisons, and the significance level was $p < .05$ for the analyses.

For most of the dependent measures, the researchers conducted a two-way ANOVA, which produced results concerning the main effects of the independent variables (i.e., equipment and taskload) and the two-way interaction between the variables. For the intelligibility and acceptability ratings, the researchers conducted a three-way ANOVA to examine background noise as a third factor. Tables will summarize the results of the analyses and report the *F* statistics associated with the effects for each dependent measure. Graphs will present the means of the experimental conditions in more detail for selected dependent measures.

3.1 System Effectiveness Measures

Table 6 shows the results of the two-way ANOVA for the system effectiveness measures. As expected, the *F* statistics indicate that controller taskload had a very strong effect on the system effectiveness measures. The safety indicators showed that controllers committed more standard and longitudinal separation errors in high taskload scenarios. The capacity indicators showed that controllers handled and completed more flights and the aircraft density was higher in high taskload scenarios. The efficiency indicators showed that controllers communicated more frequently and communicated longer in high taskload scenarios. The duration of the flights and distance flown were also longer in high taskload scenarios. However, there were no significant effects of the communication equipment and no interactions between equipment and taskload for this set of measures.

Figure 3 and Figure 4 illustrate number of push-to-talk transmissions (NPTT) and duration of push-to-talk transmissions (DPTT), respectively, as a function of the communication equipment and controller taskload. Both measures are extremely important in an equipment evaluation because any unclear pilot transmissions should result in additional controller transmissions for clarification. As shown in the figures, high taskload scenarios significantly increased NPTT and

Table 6. *F* Statistics Obtained from the Two-way ANOVA Performed on the System Effectiveness Measures

Measure	Main Effect: Equipment	Main Effect: Taskload	Interaction Effect
NSTCNF - standard conflicts	F (2, 30) = 0.18, n.s.	F (1, 15) = 8.33*	F (2, 30) = 0.50, n.s.
NLCNF - longitudinal conflicts	F (2, 30) = 1.28, n.s.	F (1, 15) = 32.17**	F (2, 30) = 0.35, n.s.
NCOMP - flights completed	F (2, 30) = 1.79, n.s.	F (1, 15) = 185.02**	F (2, 30) = 0.56, n.s.
NHAND - flights handled	F (2, 30) = 0.38, n.s.	F (1, 15) = 7418.38**	F (2, 30) = 0.10, n.s.
CMAV - aircraft density	F (2, 30) = 0.36, n.s.	F (1, 15) = 443.81**	F (2, 30) = 0.91, n.s.
NPTT - number of transmissions	F (2, 30) = 0.88, n.s.	F (1, 15) = 558.45**	F (2, 30) = 0.11, n.s.
DPTT - duration of transmissions	F (2, 30) = 0.70, n.s	F (1, 15) = 556.11**	F (2, 30) = 0.24, n.s.
NALT - altitude clearances	F (2, 30) = 2.02, n.s.	F (1, 15) = 138.87**	F (2, 30) = 1.45, n.s.
NHDG - heading clearances	F (2, 30) = 1.64, n.s.	F (1, 15) = 244.64**	F (2, 30) = 1.10, n.s.
NSPD - airspeed clearances	F (2, 30) = 0.04, n.s.	F (1, 15) = 100.23**	F (2, 30) = 0.43, n.s.
DHAND - duration of flights	F (2, 30) = 0.74, n.s.	F (1, 15) = 438.31**	F (2, 30) = 0.93, n.s.
DIST - distance of flights	F (2, 30) = 1.18, n.s.	F (1, 15) = 358.38**	F (2, 30) = 1.22, n.s.

* indicates a statistically reliable effect at a significance level of $p < .05$
** indicates a statistically reliable effect at a significance level of $p < .01$
n.s. indicates an effect that was not statistically significant

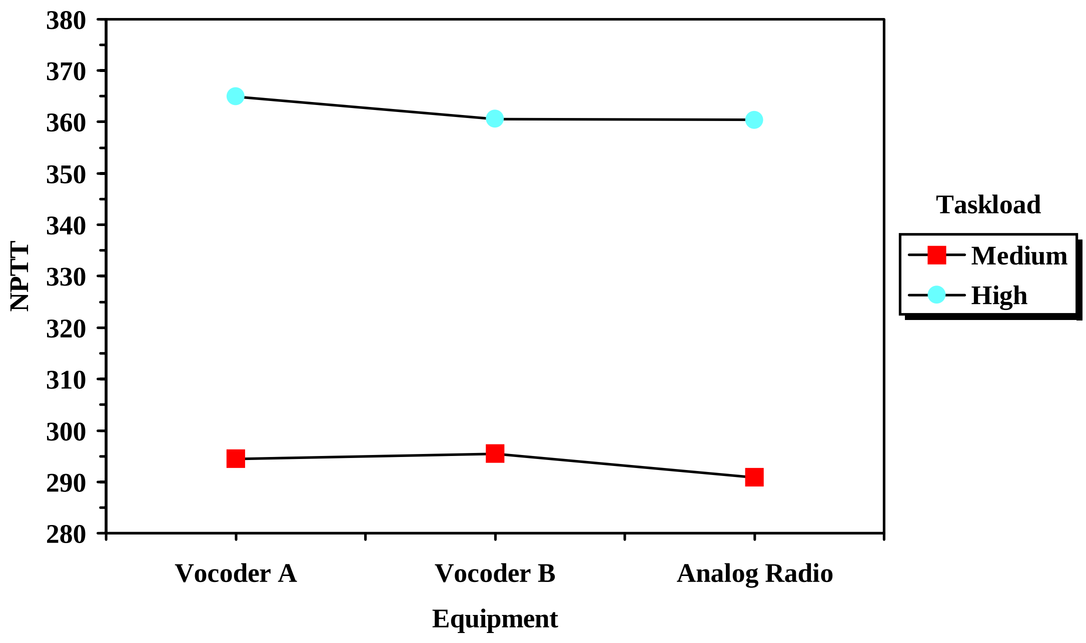

Figure 3. Mean number of push-to-talk transmissions as a function of communication equipment and controller taskload.

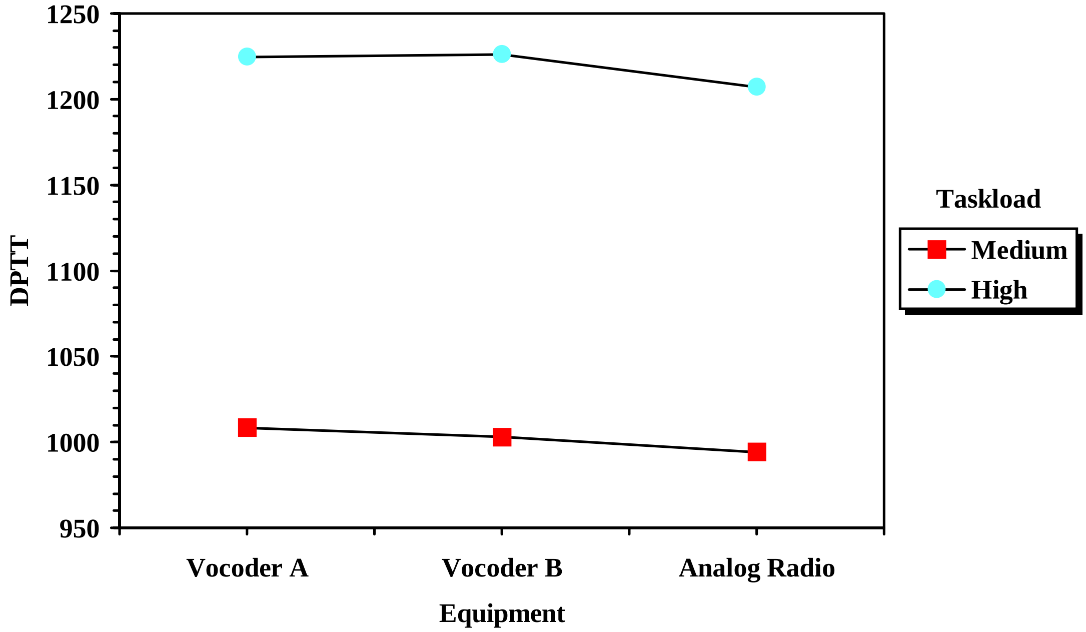

Figure 4. Mean duration of push-to-talk transmissions as a function of communication equipment and controller taskload.

DPTT. However, there were no significant effects of the communication equipment and no interactions between equipment and taskload for either measure.

3.2 Observer Ratings

Table 7 shows the results of the two-way ANOVA for the observer ratings. The F statistics indicate that controller taskload had a very strong effect on most of the observer ratings. In general, the ratings were lower in high taskload scenarios. However, taskload was not significant for observer ratings of marking flight strips, knowing LOAs and SOPs, knowing aircraft capabilities, using proper phraseology, and overall communicating. The communication equipment had no effect on the observer ratings except for listening to pilots, and there were no interactions between equipment and taskload for this set of ratings.

Table 7. *F* Statistics Obtained from the Two-way ANOVA Performed on the Observer Ratings

Rating	Main Effect: Equipment	Main Effect: Taskload	Interaction Effect
1. Maintaining separation	*F* (2, 30) = 0.02, n.s.	*F* (1, 15) = 10.07**	*F* (2, 30) = 0.05, n.s.
2. Sequencing traffic	*F* (2, 30) = 2.18, n.s.	*F* (1, 15) = 15.53**	*F* (2, 30) = 2.39, n.s.
3. Using control instructions	*F* (2, 30) = 0.35, n.s.	*F* (1, 15) = 12.79**	*F* (2, 30) = 0.55, n.s.
4. Overall traffic flow	*F* (2, 30) = 0.23, n.s.	*F* (1, 15) = 16.22**	*F* (2, 30) = 2.35, n.s.
5. Maintaining awareness	*F* (2, 30) = 0.12, n.s.	*F* (1, 15) = 15.85**	*F* (2, 30) = 1.31, n.s.
6. Ensuring positive control	*F* (2, 30) = 0.15, n.s.	*F* (1, 15) = 26.79**	*F* (2, 30) = 0.79, n.s.
7. Detecting pilot deviations	*F* (2, 30) = 0.41, n.s.	*F* (1, 15) = 9.57**	*F* (2, 30) = 0.74, n.s.
8. Correcting own errors	*F* (2, 30) = 1.84, n.s.	*F* (1, 15) = 6.55*	*F* (2, 30) = 0.30, n.s.
9. Overall attention & awareness	*F* (2, 30) = 0.13, n.s.	*F* (1, 15) = 17.87**	*F* (2, 30) = 0.61, n.s.
10. Taking action in order	*F* (2, 30) = 0.10, n.s.	*F* (1, 15) = 13.87**	*F* (2, 30) = 1.78, n.s.
11. Preplanning control actions	*F* (2, 30) = 0.25, n.s.	*F* (1, 15) = 12.33**	*F* (2, 30) = 0.78, n.s.
12. Handling control tasks	*F* (2, 30) = 0.38, n.s.	*F* (1, 15) = 16.56**	*F* (2, 30) = 1.87, n.s.
13. Marking flight strips	*F* (2, 19) = 0.61, n.s.	*F* (1, 9) = 3.77, n.s.	*F* (2, 14) = 0.00, n.s.
14. Overall prioritizing	*F* (2, 30) = 0.10, n.s.	*F* (1, 15) = 12.61**	*F* (2, 30) = 1.65, n.s.
15. Providing essential info	*F* (2, 30) = 0.82, n.s.	*F* (1, 15) = 7.35*	*F* (2, 28) = 0.53, n.s.
16. Providing additional info	*F* (2, 28) = 1.01, n.s.	*F* (1, 13) = 14.30**	*F* (2, 26) = 0.38, n.s.
17. Overall providing info	*F* (2, 30) = 1.82, n.s.	*F* (1, 15) = 10.03**	*F* (2, 29) = 1.35, n.s.
18. Knowing LOAs and SOPs	*F* (2, 30) = 0.20, n.s.	*F* (1, 15) = 3.39, n.s.	*F* (2, 29) = 0.02, n.s.
19. Knowing aircraft capabilities	*F* (2, 30) = 0.23, n.s.	*F* (1, 15) = 2.25, n.s.	*F* (2, 30) = 0.02, n.s.
20. Overall technical knowledge	*F* (2, 30) = 0.47, n.s.	*F* (1, 15) = 4.60*	*F* (2, 30) = 0.69, n.s.
21. Using proper phraseology	*F* (2, 30) = 0.74, n.s.	*F* (1, 15) = 2.81, n.s.	*F* (2, 30) = 0.03, n.s.
22. Communicating clearly	*F* (2, 30) = 0.69, n.s.	*F* (1, 15) = 4.62*	*F* (2, 30) = 0.40, n.s.
23. Listening to pilots	*F* (2, 30) = 3.33*	*F* (1, 15) = 8.80**	*F* (2, 30) = 0.45, n.s.
24. Overall communicating	*F* (2, 30) = 1.08, n.s.	*F* (1, 15) = 3.00, n.s.	*F* (2, 30) = 0.38, n.s.

* indicates a statistically reliable effect at a significance level of $p < .05$
** indicates a statistically reliable effect at a significance level of $p < .01$
n.s. indicates an effect that was not statistically significant

Figure 5 illustrates the observer ratings for listening to pilots as a function of the communication equipment and controller taskload. Although the difference appears small, observer ratings were significantly lower in high taskload scenarios. Because the equipment effect was significant also, the researchers conducted Tukey HSD post hoc comparisons. The tests revealed that vocoder A received the highest observer ratings and there was no significant difference between analog radio and vocoder B.

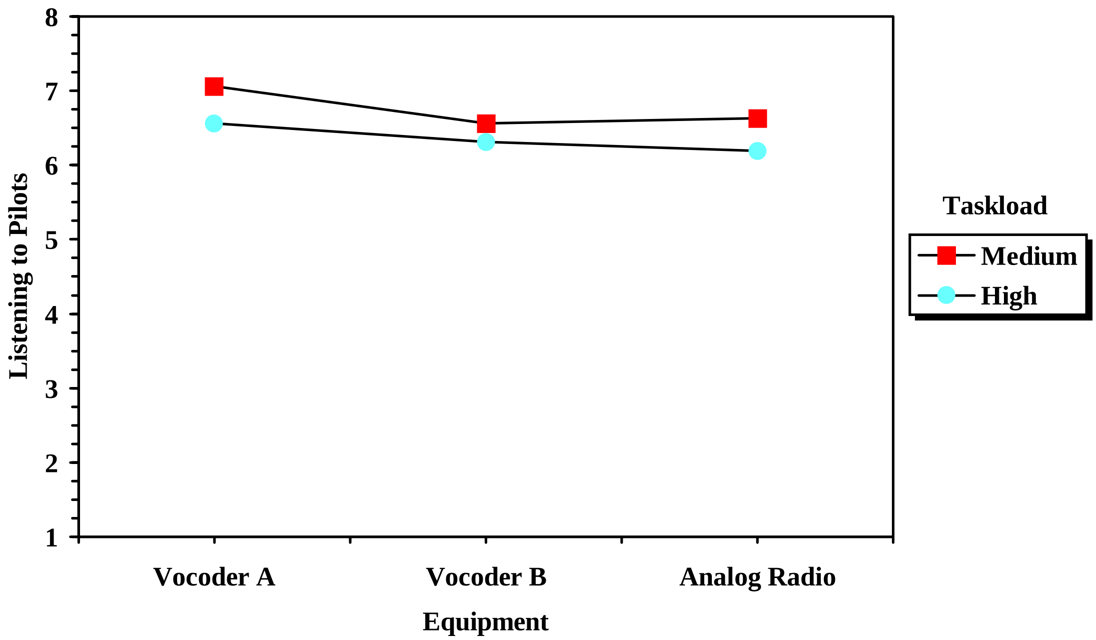

Figure 5. Mean observer rating for listening to pilot readbacks and requests as a function of communication equipment and controller taskload.

Figure 6 illustrates a taxonomy of the observer comments recorded during the simulation. The purpose of the taxonomy was to identify any differences in controller performance using the three communication systems. The researchers selected 23 categories based upon a subjective determination of common themes within the observer comments. The researchers computed the percentages for each communication system based upon 411 comments for vocoder A, 450 comments for vocoder B, and 445 comments for analog radio. Although the researchers did not conduct any formal statistical procedures on the taxonomy, there do not appear to be any large differences between the communication systems. As shown, the most frequent observer comment referred to excessive final spacing.

3.3 Controller Ratings

Table 8 shows the results of the two-way ANOVA for the controller ratings. The *F* statistics indicate that controller taskload had a very strong effect on most of the controller ratings. Controller and simulation pilot performance was lower in high taskload scenarios. Mental, physical, temporal, and overall workload were higher in high taskload scenarios. Controller effort and frustration were also higher in high taskload scenarios. However, taskload was not significant for situation awareness ratings and overall intelligibility and acceptability ratings. The communication equipment had a significant effect on overall intelligibility and acceptability ratings, but there were no interactions between equipment and taskload for this set of ratings.

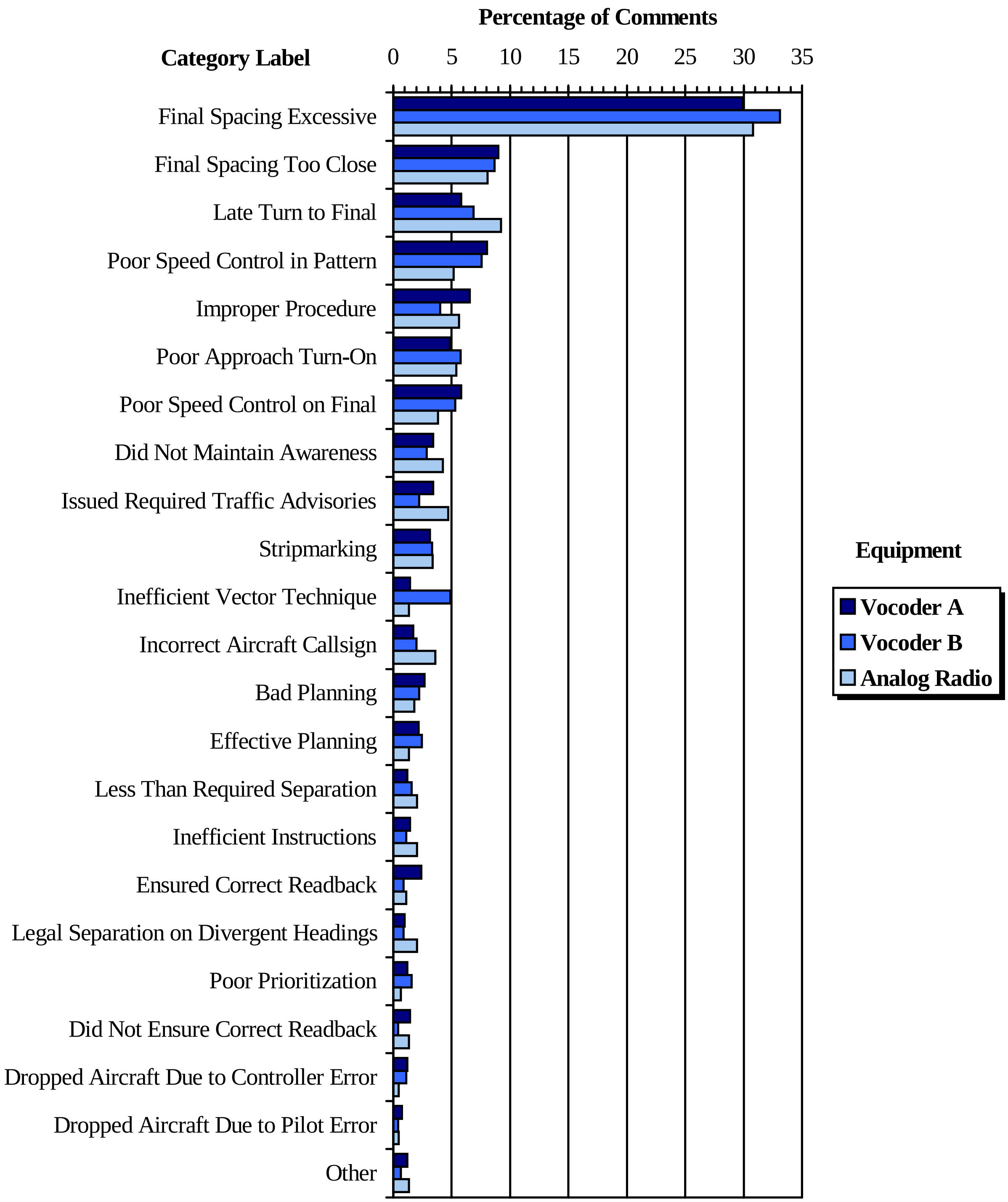

Figure 6. Taxonomy of observer comments as a function of the communication equipment.

Table 8. *F* Statistics Obtained from the Two-way ANOVA Performed on the Controller Ratings

Rating	Main Effect: Equipment	Main Effect: Taskload	Interaction Effect
1. Controller performance	F (2, 30) = 0.93, n.s.	F (1, 15) = 15.92**	F (2, 30) = 0.17, n.s.
2. Controller workload	F (2, 30) = 0.79, n.s.	F (1, 15) = 256.58**	F (2, 30) = 0.37, n.s.
3. Controller situation awareness	F (2, 30) = 1.11, n.s.	F (1, 15) = 2.72, n.s.	F (2, 30) = 0.09, n.s.
4. Simulation pilot performance	F (2, 30) = 0.06, n.s.	F (1, 15) = 9.40**	F (2, 30) = 1.33, n.s.
5. NASA-TLX, mental demand	F (2, 30) = 0.01, n.s.	F (1, 15) = 157.08**	F (2, 30) = 2.93, n.s.
6. NASA-TLX, physical demand	F (2, 30) = 0.25, n.s.	F (1, 15) = 70.00**	F (2, 30) = 0.73, n.s.
7. NASA-TLX, temporal demand	F (2, 30) = 0.69, n.s.	F (1, 15) = 136.13**	F (2, 30) = 0.46, n.s.
8. NASA-TLX, performance	F (2, 30) = 0.42, n.s.	F (1, 15) = 7.27*	F (2, 30) = 0.21, n.s.
9. NASA-TLX, effort	F (2, 30) = 0.48, n.s.	F (1, 15) = 16.65**	F (2, 30) = 0.26, n.s.
10. NASA-TLX, frustration	F (2, 30) = 0.23, n.s.	F (1, 15) = 23.43**	F (2, 30) = 0.00, n.s.
11a. Intelligibility, overall	F (2, 30) = 10.21**	F (1, 15) = 0.45, n.s.	F (2, 30) = 0.89, n.s.
11b. Acceptability, overall	F (2, 30) = 16.54**	F (1, 15) = 0.20, n.s.	F (2, 30) = 1.31, n.s.
ATWIT	F (2, 30) = 2.24, n.s.	F (1, 15) = 119.01**	F (2, 30) = 0.13, n.s.

* indicates a statistically reliable effect at a significance level of $p < .05$
** indicates a statistically reliable effect at a significance level of $p < .01$
n.s. indicates an effect that was not statistically significant

Figure 7 illustrates the ATWIT ratings as a function of the communication equipment and controller taskload. Controller workload is an important measure in an equipment evaluation because any difficulty in communications should result in higher workload ratings. As shown in the figure, high taskload scenarios significantly increased workload, but equipment had no effect.

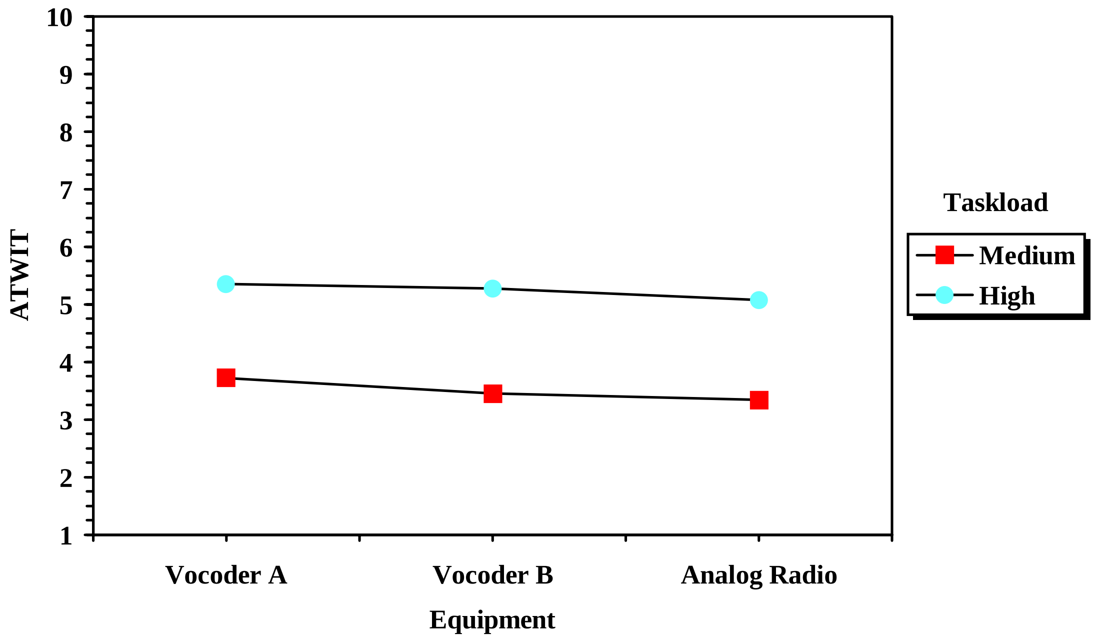

Figure 7. Mean Air Traffic Workload Input Technique ratings as a function of communication equipment and controller taskload.

Figure 8 and Figure 9 illustrate the intelligibility and acceptability ratings, respectively, for all transmissions as a function of the communication equipment and controller taskload. The patterns of the ratings were nearly identical, although intelligibility ratings were slightly higher than acceptability ratings. In fact, the Pearson product-moment correlation between the intelligibility and acceptability was very high, r (190) = .88. Taskload had no effect on intelligibility and acceptability ratings. However, because the equipment effect was significant, researchers conducted Tukey HSD post hoc comparisons. The tests revealed that vocoder A was the least intelligible and least acceptable. Analog radio and vocoder B were not significantly different for either rating.

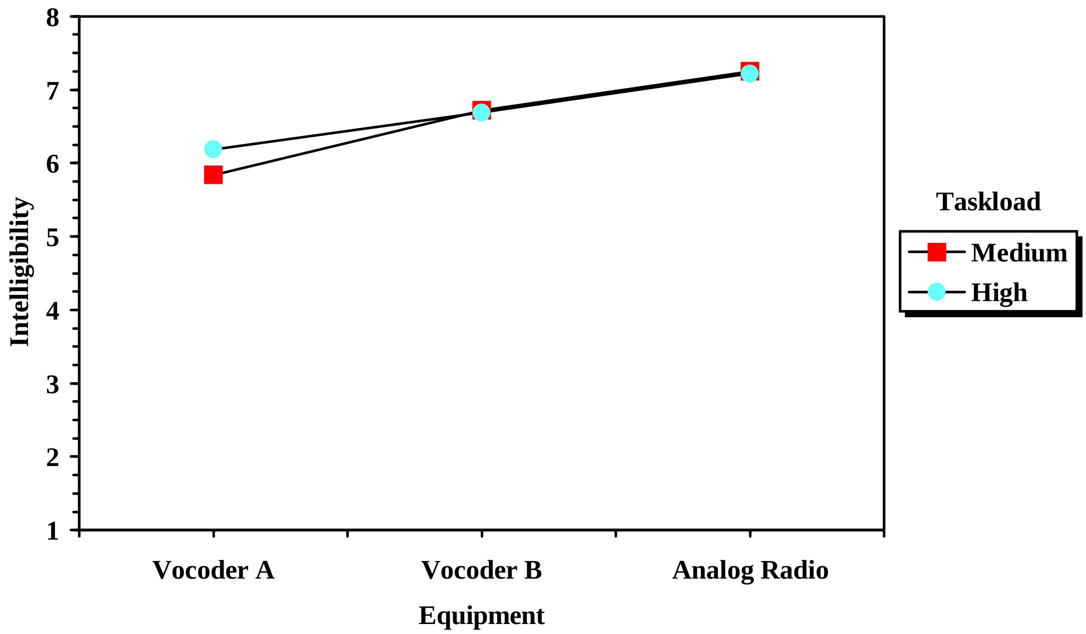

Figure 8. Mean intelligibility ratings for all transmissions as a function of communication equipment and controller taskload.

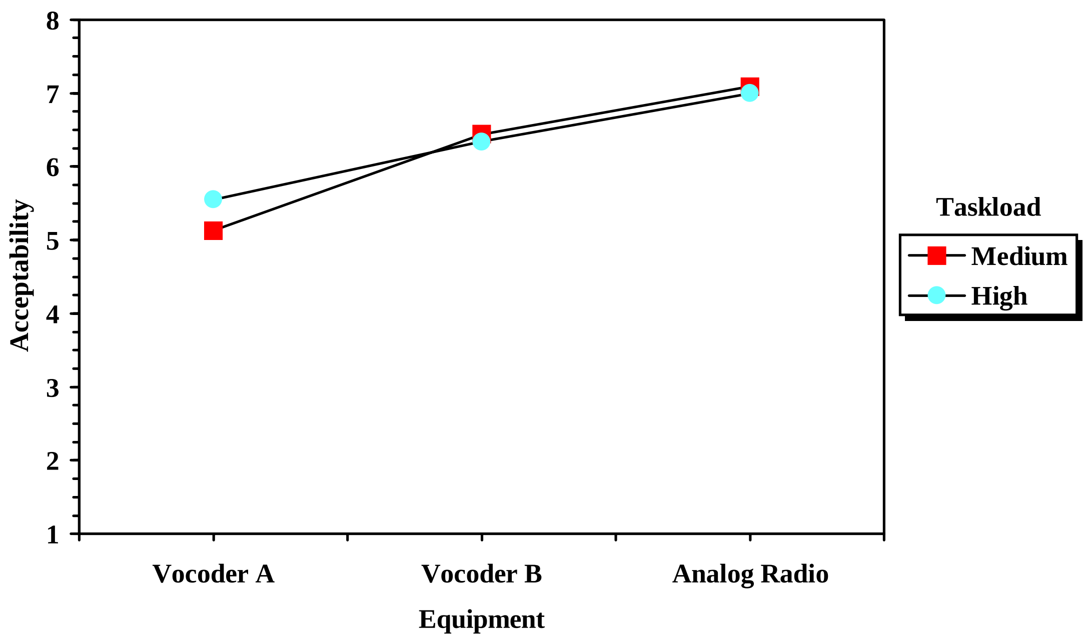

Figure 9. Mean acceptability ratings for all transmissioins as a function of communication equipment and controller taskload.

Table 9 shows the results of the three-way ANOVA performed on the intelligibility ratings with aircraft background noise as the third factor. As in the previous two-way analysis of overall intelligibility, the *F* statistics indicate that controller taskload had no effect on intelligibility ratings. The main effects of equipment and background were significant. However, the interaction between equipment and background was significant also and qualified the individual main effects. The researchers examined the simple main effects for each of the three background noises.

Table 9. Degrees of Freedom, Mean Squares, and *F* Statistics Obtained from the Three-way ANOVA Performed on the Intelligibility Ratings

Source of Variation	Degrees of Freedom	Mean Square	*F* Statistic
Equipment	2, 30	72.18	10.17**
Taskload	1, 15	2.12	0.61, n.s.
Background	2, 30	38.61	11.79**
Equipment*Taskload	2, 30	3.44	1.09, n.s.
Equipment*Background	4, 60	2.12	2.64*
Taskload*Background	2, 30	0.49	0.76, n.s.
Equipment*Taskload*Background	4, 60	0.19	0.45, n.s.

* indicates a statistically reliable effect at a significance level of $p < .05$
** indicates a statistically reliable effect at a significance level of $p < .01$
n.s. indicates an effect that was not statistically significant

Table 10 shows the results of the analysis of simple main effects and the Tukey HSD post hoc comparisons conducted on the significant effects. The *F* statistics indicate that all three simple main effects were significant. For jet and propeller background noises, vocoder A was the least intelligible and analog radio and vocoder B were not significantly different. For helicopter background noise, analog radio was the most intelligible and vocoder A and vocoder B were not significantly different.

Table 10. Mean Intelligibility Ratings, *F* Statistics Obtained from the Analysis of Simple Main Effects, and Tukey HSD Post Hoc Comparisons

For Jet Background Noises				
Vocoder A	Vocoder B	Analog Radio	*F* Statistic	Tukey HSD Comparisons
6.22	6.91	7.17	6.46**	A < B; A < Radio; B = Radio
For Propeller Background Noises				
Vocoder A	Vocoder B	Analog Radio	*F* Statistic	Tukey HSD Comparisons
5.86	6.58	7.13	11.59**	A < B; A < Radio; B = Radio
For Helicopter Background Noises				
Vocoder A	Vocoder B	Analog Radio	*F* Statistic	Tukey HSD Comparisons
5.23	5.75	6.70	8.42**	A = B; A < Radio; B < Radio

** indicates a statistically reliable effect at a significance level of $p < .01$

Table 11 shows the results of the three-way ANOVA performed on the acceptability ratings with aircraft background noise as the third factor. As in the previous two-way analysis of overall acceptability, the *F* statistics indicate that controller taskload had no effect on acceptability ratings. The main effects of equipment and background were significant. Although the interaction between equipment and background was not significant, the effect was nearly significant. Because of the importance of acceptability ratings in this study, the researchers further investigated the relationship between equipment and background by examining the simple main effects for each of the three background noises.

Table 11. Degrees of Freedom, Mean Squares, and *F* Statistics Obtained from the Three-way ANOVA Performed on the Acceptability Ratings

Source of Variation	Degrees of Freedom	Mean Square	*F* Statistic
Equipment	2, 30	106.72	12.57**
Taskload	1, 15	0.56	0.10, n.s.
Background	2, 30	43.22	10.54**
Equipment*Taskload	2, 30	2.66	0.65, n.s.
Equipment*Background	4, 60	1.89	2.40†
Taskload*Background	2, 30	0.20	0.34, n.s.
Equipment*Taskload*Background	4, 60	0.25	0.47, n.s.

** indicates a statistically reliable effect at a significance level of $p < .01$
n.s. indicates an effect that was not statistically significant
Note.
† indicates an effect that was not statistically significant, but nearly significant with a *p* value less than .06

Table 12 shows the results of the analysis of simple main effects and the Tukey HSD post hoc comparisons conducted on the significant effects. The *F* statistics indicate that all three simple main effects were significant and the pattern was the same as the intelligibility ratings. For jet and propeller background noises, vocoder A was the least acceptable and analog radio and vocoder B were not significantly different. For helicopter background noise, analog radio was the most acceptable and vocoder A and vocoder B were not significantly different.

Table 12. Mean Acceptability Ratings, *F* Statistics Obtained from the Analysis of Simple Main Effects, and Tukey HSD Post Hoc Comparisons

For Jet Background Noises				
Vocoder A	Vocoder B	Analog Radio	*F* Statistic	Tukey HSD Comparisons
5.69	6.52	6.92	8.78**	A < B; A < Radio; B = Radio
For Propeller Background Noises				
Vocoder A	Vocoder B	Analog Radio	*F* Statistic	Tukey HSD Comparisons
5.30	6.16	6.86	13.18**	A < B; A < Radio; B = Radio
For Helicopter Background Noises				
Vocoder A	Vocoder B	Analog Radio	*F* Statistic	Tukey HSD Comparisons
4.67	5.31	6.38	10.66**	A = B; A < Radio; B < Radio

** indicates a statistically reliable effect at a significance level of $p < .01$

3.4 Final Questionnaire

Table 13 shows the controller responses to questions on the final questionnaire. The results are means based upon a 10-point rating scale. As shown, controllers found the simulation to be realistic and the generic airspace easy to learn. The participants also indicated that the simulation pilots performed well and the ATWIT procedure did not interfere with their performance.

Table 13. Exit Questionnaire Ratings

Question	Mean	*SD*
1. In general, how realistic was the simulation?	6.94	2.08
2. How realistic were the aircraft background noises?	7.38	2.00
3. How realistic were the traffic scenarios?	8.13	1.73
4. How realistic was GENERA airspace?	7.69	1.62
5. How difficult was it to learn the GENERA airspace?	1.38	1.02
6. How well did the simulation pilots perform in the simulation?	7.94	1.39
7. To what extent did the ATWIT probe technique interfere with your performance?	1.88	1.26

4. Discussion and Conclusions

The communication equipment had no effect on the system effectiveness measures. Controllers maintained safety, capacity, and efficiency while using the vocoders. In general, there were few separation errors, and capacity remained constant because controllers did not hold traffic.

However, NPTT and DPTT were sensitive indicators that tended to vary with individual controller style. Even so, transmissions were no more frequent or longer using vocoders compared to analog radio.

Controller taskload had large effects on the system effectiveness measures. Safety and efficiency decreased, and capacity increased in high taskload scenarios. However, because there were no interactions between equipment and taskload, the vocoders did not impede performance in either low or high taskload scenarios. Objectively, the system effectiveness measures indicate that vocoder transmissions were highly intelligible and did not disrupt controller performance. These results are consistent with the objective intelligibility findings of the phase I study.

The observer ratings of controller performance also tended to vary with individual controller style. Although some controllers performed better than others, observer ratings were not any lower while using the vocoders. In fact, observers rated listening to pilots as higher for vocoder A than analog radio or vocoder B. The higher observer rating in this performance area was unusual because controllers tended to rate vocoder A as the least intelligible and acceptable. However, the result suggests that controllers were listening more closely to vocoder A transmissions, possibly due to a poorer quality signal, and made more readback corrections or clarifications. The subjective observer ratings were consistent with the objective system effectiveness measures, and both indicate that the vocoders did not interfere with controller performance.

Although the intelligibility and acceptability results were very similar, the correlation between ratings was much lower in the first phase ($r = .37$) compared to the second phase ($r = .88$). The reason for this difference is not clear, but it is likely due to the differences in the rating procedures. In phase I, controllers listened to audio recordings and made intelligibility and acceptability ratings immediately after the researchers presented each message. This procedure did not involve memory and seemed to encourage controllers to contrast intelligibility and acceptability and make independent ratings. In phase II, controllers made post-scenario ratings that depended upon memory and seemed to encourage related intelligibility and acceptability ratings.

The results of both phases showed that the signal quality of the vocoders was different for the three aircraft background noises. For jet and propeller background noises, vocoder B was as intelligible and acceptable as analog radio, but vocoder A was slightly lower. In fact, both vocoders had some difficulty processing helicopter background noises compared to analog radio. The reason for these differences is likely due to the different speech models and compression algorithms of the vocoders. The speech model for vocoder B seemed to be more effective than vocoder A, although helicopter background noise was a weakness for both. Now that this study has identified these weaknesses, it may be possible for the vocoder manufacturers to improve upon their models in future versions.

The present research demonstrates the power of simulation to evaluate new concepts and equipment. Simulation places controllers under realistic taskloads and demands performance under conditions that they have experienced in their facilities. Simulation allows researchers to make empirical comparisons of current technology with advanced systems or subsystems. This study demonstrates the capabilities of simulation to go beyond subjective analyses and provide managers with objective performance data to make decisions about proposed changes to the ATC system.

The results of both phases showed that intelligibility and acceptability ratings were very high and nearly equal for analog radio and vocoder B and only slightly lower for vocoder A. These results, coupled with the lack of any performance differences using the vocoders, suggest that vocoder technology could replace the current analog radio system in the future. However, both phases of the study have examined a limited set of factors that could potentially influence the effectiveness of vocoders. Future research should address other issues such as the effects of speech rate, accents, pilot reception of controller transmissions, and signal degradation over distance.

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-TN83/26). Atlantic City, NJ: DOT/FAA Technical Center.

Child, J., Cleve, R., & Grable, M. (1989). *Evaluation of low data rate CODECS for air traffic control applications* (DOT/FAA/CT-TN89/13). Atlantic City, 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, NJ: DOT/FAA 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, NJ: DOT/FAA Technical Center.

Hart, S. G., & Staveland, L. E. (1988). 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 Press.

La Due, J. C., Sollenberger, R. L., Belanger, B., & Heinze, A. (1997) *Human factors evaluation of vocoders for ATC environments: phase I report* (DOT/FAA/CT-TN97/11). Atlantic City, NJ: DOT/FAA William J. Hughes Technical Center.

Sollenberger, R. L., Stein, E. S., & Gromelski, S. (1997). *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, 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, 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). *Simulation variables*. Unpublished manuscript.

UFA, Inc. (1992). *ATCoach* [Computer software]. Lexington, MA: UFA, Inc.

Appendix A
ATC System Effectiveness Measures

I – Safety Indicators
NSTCNF - Number of standard terminal conflicts
DSTCNF - Duration of standard terminal conflicts
NTCNF - Number of user-defined terminal conflicts
DTCNF - Duration of user-defined terminal conflicts
NLCNF - Number of ILS conflicts
DLCNF - Duration of ILS conflicts
NPCNF - Number of parallel conflicts
NBSCNF - Number of between sector conflicts
DBSCNF - Duration of between sector conflicts
NASCNF - Number of airspace violations
DASCNF - Duration of airspace violations
API - Aircraft proximity index
CPA - Closest point of approach for each conflict
CPAHSEP - Horizontal separation at CPA time
CPAVSEP - Vertical separation at CPA time
NHOMISS - Number of handoff misses

II – Capacity Indicators
CMAV – Cumulative average of system activity
NHAND – Number of flights handled
NCOMP – Number of flights completed
NLAND – Number of arrivals completed
NDEP – Number of departures completed
NHOFF – Number of successful handoffs

III – Efficiency Indicators
NPTT - Number of controller push-to-talk transmissions DPTT - Duration of controller push-to-talk transmissions NALT - Number of altitude clearances NHDG - Number of heading clearances NSPD - Number of airspeed clearances DHAND - Duration of flights handled AVLAND - Average landing interval time AVDEP - Average departure interval time DHODLY - Duration of handoff delays NHTDLY - Number of hold/turn delays DHTDLY - Duration of hold/turn delays NSTDLY - Number of start point delays DSTDLY - Duration of start point delays NMISS - Number of missed approaches NCMESG - Number of controller key/slew entries

Appendix B
Observer Rating Form

Observer Code _________ Date _________

Participant: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Scenario: M1 M2 M3 M4 M5 M6 H1 H2 H3 H4 H5 H6

Equipment: A B Radio

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. Your identity will remain anonymous, as your data will be identified by an observer code known only to yourself and the researchers conducting this study. 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.

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

1. Maintaining Separation and Resolving Potential Conflicts **1 2 3 4 5 6 7 8**
   - using control instructions that maintain safe aircraft separation
   - detecting and resolving impending conflicts early
   - recognizing the need for speed restrictions and wake turbulence separation

Comments:

2. Sequencing Arrival and Departure Aircraft Efficiently.................... **1 2 3 4 5 6 7 8**
   - using efficient and orderly spacing techniques for arrival and departure aircraft
   - maintaining safe arrival and departure intervals that minimize delays

Comments:

3. Using Control Instructions Effectively/Efficiently........................... **1 2 3 4 5 6 7 8**
   - providing accurate navigational assistance to pilots
   - issuing economical clearances that result in need for few additional instructions to handle aircraft completely
   - ensuring clearances use minimum necessary flight path changes

Comments:

4. Overall Safe and Efficient Traffic Flow Scale Rating...................... **1 2 3 4 5 6 7 8**

Comments:

II - MAINTAINING ATTENTION AND SITUATION AWARENESS

5. Maintaining Awareness of Aircraft Positions **1 2 3 4 5 6 7 8**
   - 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

Comments:

6. Ensuring Positive Control **1 2 3 4 5 6 7 8**
   - tailoring control actions to situation
   - using standard procedures for handling heavy, emergency, and unusual traffic situations
   - ensuring pilot adherence to issued clearances

Comments:

7. Detecting Pilot Deviations from Control Instructions **1 2 3 4 5 6 7 8**
   - ensuring that pilots follow assigned clearances correctly
   - correcting pilot deviations in a timely manner

Comments:

8. Correcting Own Errors in a Timely Manner **1 2 3 4 5 6 7 8**
   - acting quickly to correct errors
   - changing an issued clearance when necessary to expedite traffic flow

Comments:

9. Overall Attention and Situation Awareness Scale Rating **1 2 3 4 5 6 7 8**

Comments:

III - PRIORITIZING

10. Taking Actions in an Appropriate Order of Importance.................. **1 2 3 4 5 6 7 8**
    - resolving situations that need immediate attention before handling low priority tasks
    - issuing control instructions in a prioritized, structured, and timely manner

Comments:

11. Preplanning Control Actions... **1 2 3 4 5 6 7 8**
    - scanning adjacent sectors to plan for future and conflicting traffic
    - studying pending flight strips in bay

Comments:

12. Handling Control Tasks for Several Aircraft **1 2 3 4 5 6 7 8**
    - shifting control tasks between several aircraft when necessary
    - communicating in timely fashion while sharing time with other actions

Comments:

13. Marking Flight Strips while Performing Other Tasks...................... **1 2 3 4 5 6 7 8**
    - marking flight strips accurately while talking or performing other tasks
    - keeping flight strips current

Comments:

14. Overall Prioritizing Scale Rating **1 2 3 4 5 6 7 8**

Comments:

IV - Providing Control Information

15. Providing Essential Air Traffic Control Information **1 2 3 4 5 6 7 8**
    - providing mandatory services and advisories to pilots in a timely manner
    - exchanging essential information

Comments:

16. Providing Additional Air Traffic Control Information..................... **1 2 3 4 5 6 7 8**
    - providing additional services when workload is not a factor
    - exchanging additional information

Comments:

17. Overall Providing Control Information Scale Rating.......................**1 2 3 4 5 6 7 8**

Comments:

V - Technical Knowledge

18. Showing Knowledge of LOAs and SOPs **1 2 3 4 5 6 7 8**
    - controlling traffic as depicted in current LOAs and SOPs
    - performing handoff procedures correctly

Comments:

19. Showing Knowledge of Aircraft Capabilities and Limitations **1 2 3 4 5 6 7 8**
    - using appropriate speed, vectoring, and/or altitude assignments to separate aircraft with varied flight capabilities
    - issuing clearances that are within aircraft performance parameters

Comments:

20. Overall Technical Knowledge Scale Rating **1 2 3 4 5 6 7 8**

Comments:

VI - COMMUNICATING

21. Using Proper Phraseology.. **1 2 3 4 5 6 7 8**
    - using words and phrases specified in the 7110.65
    - using phraseology that is appropriate for the situation
    - using minimum necessary verbiage
    - speaking with confident, authoritative tone of voice

Comments:

22. Communicating Clearly and Efficiently... **1 2 3 4 5 6 7 8**
    - 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
    - providing complete information in each clearance

Comments:

23. Listening to Pilot Readbacks and Requests..................................... **1 2 3 4 5 6 7 8**
    - correcting pilot readback errors
    - acknowledging pilot or other controller requests promptly
    - processing requests correctly in a timely manner

Comments:

24. Overall Communicating Scale Rating **1 2 3 4 5 6 7 8**

Comments

Appendix C
Post-Scenario Questionnaire

Participant: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Scenario: Mx M1 M2 M3 M4 M5 M6 Hx H1 H2 H3 H4 H5 H6
Equipment: A B Radio

INSTRUCTIONS

The purpose of this questionnaire is to determine how the conditions of this scenario affect your opinions and performance. As you answer each question, please be as honest and as accurate as you can. Your identity will remain anonymous, so do not write your name on the form. Instead, your data will be identified by a participant code known only to yourself and the researchers conducting this study.

General Ratings

1. Please rate **how well you controlled traffic** during this scenario.

not well	1	2	3	4	5	6	7	8	9	10	extremely well

2. Please rate your **overall workload** during this scenario.

very low	1	2	3	4	5	6	7	8	9	10	very high

3. Please rate your **overall situational awareness** during this scenario.

very low	1	2	3	4	5	6	7	8	9	10	very high

4. Please rate **how well the simulation pilots performed** during this scenario.

not well	1	2	3	4	5	6	7	8	9	10	extremely well

NASA TLX

5. Circle the number that best describes the **mental demand** during this scenario.

extremely low	1	2	3	4	5	6	7	8	9	10	extremely high

6. Circle the number that best describes the **physical demand** during this scenario.

extremely low	1	2	3	4	5	6	7	8	9	10	extremely high

7. Circle the number that best describes the **temporal demand** during this scenario.

extremely low	1	2	3	4	5	6	7	8	9	10	extremely high

8. Circle the number that best describes your **performance** during this scenario.

extremely low	1	2	3	4	5	6	7	8	9	10	extremely high

9. Circle the number that best describes your **effort** during this scenario.

extremely low	1	2	3	4	5	6	7	8	9	10	extremely high

10. Circle the number that best describes your **level of frustration** during this scenario.

extremely low	1	2	3	4	5	6	7	8	9	10	extremely high

INSTRUCTIONS

In the scenario just completed, transmissions from the simulation pilots have been processed through either a vocoder or an analog radio simulator. Please rate the intelligibility and the acceptability of the pilot transmissions on the scales defined below. Confine your ratings to the scenario just completed. Circle the one number that best applies for each scale.

Intelligibility

- Ability to understand what was said in the message

poor	1	2	3	4	5	6	7	8	excellent

Poor - could not understand anything that was said during the transmission

Excellent - understood everything that was relayed during the transmission precisely

Acceptability

- Quality of the message: e.g., annoying, pleasant
- Effort required to understand the message: e.g., easy, burdensome
- Potential influence of the background noise: e.g., buzzing, hissing, etc.

poor	1	2	3	4	5	6	7	8	excellent

Poor - terribly annoying, frustrating, or unpleasant to listen to

Excellent - excellent signal quality, a clear signal that was pleasant to listen to

Intelligibility

Poor - could not understand anything that was said during the transmission
Excellent - understood everything that was relayed during the transmission precisely

Acceptability

Poor - terribly annoying, frustrating, or unpleasant to listen to
Excellent - excellent signal quality, a clear signal that was pleasant to listen to

---

11. **In general, all** transmissions

Intelligibility

poor	1	2	3	4	5	6	7	8	excellent

Acceptability

poor	1	2	3	4	5	6	7	8	excellent

12. **Jet background** transmissions

Intelligibility

poor	1	2	3	4	5	6	7	8	excellent

Acceptability

poor	1	2	3	4	5	6	7	8	excellent

13. **Propeller background** transmissions

Intelligibility

poor	1	2	3	4	5	6	7	8	excellent

Acceptability

poor	1	2	3	4	5	6	7	8	excellent

14. **Helicopter background** transmissions

Intelligibility

poor	1	2	3	4	5	6	7	8	excellent

Acceptability

poor	1	2	3	4	5	6	7	8	excellent

Please take a moment and briefly write some notes about your impressions of the scenario just completed. Focus on the communications and any problems you might have encountered. Be as specific as you can.