HUMAN FACTORS EVALUATION OF A DIGITAL, AIR-GROUND COMMUNICATIONS SYSTEM

Carolina M. Zingale, Ph.D., Titan Corporation, Mays Landing, New Jersey
D. Michael McAnulty, Ph.D., FAA William J. Hughes Technical Center,
Atlantic City International Airport, New Jersey
Karol Kerns, Ph.D., MITRE, McLean, Virginia

Abstract

The Federal Aviation Administration (FAA) Next Generation Air-Ground Communications (NEXCOM) program has been developing a Very High Frequency (VHF) Digital Link Mode 3 (VDL3) system to replace the aging analog air traffic communications system. VDL3 provides increased channel capacity and is capable of transmitting both voice and data. VDL3 also compensates for known limitations in the analog system by virtually eliminating "step-ons" with an antiblocking feature, enabling controller override, and providing a transmit status indicator (TSI) to indicate if the channel is occupied.

VDL3 is also expected to have a longer voice throughput delay (up to 350 ms) than the existing analog system (approximately 70 ms), which could potentially disrupt the communications flow. The delay and the acceptability of the new features were previously evaluated in a high-fidelity human-in-the-loop simulation with air traffic control specialists (ATCSs) [1]. That study found that a system with a 350 ms system delay and the additional features was acceptable to controllers and did not adversely affect performance or workload. This report summarizes the second, high fidelity, human-in-the-loop simulation of VDL3 system performance and operational acceptability from the flight deck perspective. The objectives of this study were to validate the findings of the earlier simulation with pilots, to compare data obtained under analog communications to those obtained using a digital system simulating VDL3, and to assess analog and VDL3 communications under routine conditions and adverse weather conditions that further increased demand for access to the channel.

Fourteen airline pilots participated in the study using two realistic flight deck simulators. The results showed that the participants attempted to access the channel similarly with either radio system, but that the digital system allowed more successful transmissions to be made. Most other communications characteristics did not differ between the two systems. The effects of adverse weather were similar for both systems. The participants rated the operational acceptability of the digital system higher than the analog system, and nearly always rated the digital system as equal to or better than the analog system for completing communication tasks. The participants rated the antiblocking, controller override, and TSI features as highly useful. However, ratings of some aspects of the TSI were variable, indicating that improvements may be needed. In a separate effort, a group of human factors specialists evaluated alternative implementations of the TSI and made recommendations for modifications. Overall, the results indicated that VDL3, with a 350 ms voice throughput delay and enhanced system features, is an acceptable communications system for pilots.

Introduction

The Federal Aviation Administration (FAA) Next Generation Air-Ground Communications (NEXCOM) program has been developing a Very High Frequency (VHF) Digital Link Mode 3 (VDL3) system to replace the current analog air traffic communications system. The current system is decades old, and much of the infrastructure is at the end of its service life. There are also several problems with the system, such as "step-ons" that occur when two or more users try to transmit simultaneously and in which none of the messages is sent clearly and completely. When the communication frequency is very busy or is blocked by a stuck microphone, the air traffic controller (who makes almost half the total transmissions) cannot override the system to send urgent messages. In a step-on situation, the parties attempting to transmit are unaware that their messages have not been sent clearly (they only hear themselves transmitting). Finally, the increasing volume of air traffic is straining the capacity of the available frequency spectrum for analog communications using a 25 kHz bandwidth. The European states have addressed their spectrum problem by reducing the bandwidth to 8.33 kHz, but the U.S. views this as only an interim solution.

VDL3 is a time division, multiple access system that provides increased channel capacity (i.e., four time slots per 25 kHz frequency) and is capable of transmitting both voice and data. The VDL3 system compensates for known limitations in the analog system by implementing features such as controller override, antiblocking, and a transmit status indicator (TSI). These features are designed to allow for more efficient channel access and provide notification about channel availability. This system virtually eliminates "step-ons." However, the proposed system will also have a longer voice throughput delay (up to 350 ms) than the analog system (approximately 70 ms).

Purpose

An earlier simulation evaluated the effects of three VDL3 system voice throughput delays (250, 350, and 750 ms) and the antiblocking, TSI, and controller override features on air traffic control specialist (ATCS) performance and workload [1]. That study found that a system with a 350 ms ground system delay with the additional features would be effective and acceptable to controllers. The current study was designed to validate those findings with pilots, to compare data obtained under analog communications to those obtained using VDL3, and to assess communications under routine conditions and under adverse weather conditions that would further increase user demand for the channel.

Methods

Participants

Fourteen certified airline pilots served as participants. They were recruited through the Air Line Pilots Association (ALPA). Twelve participants were current for Instrument Flight Rules (IFR). The other two were furloughed. The participants ranged in age from 23 to 53 with a mean age of 39. Thirteen of them were male. The participants reported an average of nearly 8,000 total flight hours, including military experience. All indicated having previous experience with simulators. They rated their satisfaction with the current communications system between 2 and 8, with an average rating of 5.1 on a 10-point scale in which 1 represented the lowest rating and 10 the highest rating.

Test Facility

We used two realistic flight deck simulators in the study. One simulator was a fixed-base, transport aircraft with a two-pilot crew that emulated a Boeing 747-400 (B-747). The other simulator was a motion-based, twin engine general aviation cockpit that emulated a Cessna 421 (C-421). The cockpit environments included jet and propeller engine noise, respectively, and the C-421 could simulate turbulence.

We linked the two simulators and 12 additional pseudopilot workstations to the Target Generator Facility (TGF). The TGF controlled the aircraft maneuvers based on simulation pilot and pseudopilot entries and on scripted flight plan data. These entries and data were then sent as input to the controller simulation workstation, the Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE). DESIREE emulates Display System Replacement (DSR) functionality and includes a high-resolution Sony 2K display monitor, en route keyboard and trackball, headsets with microphones, and push-to-talk (PTT) handsets or foot pedals. For our simulation, the workstation was configured for single controller operation.

To simulate the communications systems, we used a Yamaha D5000 Digital Delay System to implement system delays. To simulate current analog voice communications, the system provided an audio delay of approximately 70 ms ground-to-air and air-to-ground, with no delay between the pseudopilot positions. To simulate the VDL3 system, we implemented a 350 ms ground delay. The total transmission propagation and avionics processing delays resulted in an end-to-end, ground-to-air communication delay of 390 ms and an air-to-ground delay of 360 ms. We further modified the communications system to implement the controller override, antiblocking, and TSI features exactly as they had been in the earlier evaluation [1]. In this configuration, controller override occurred immediately when the controller pressed the PTT key during a pilot transmission[1]. The TSI consisted of an audio signal to the pilot that occurred within 500 ms after the pilot attempted to transmit on an already-occupied channel. This indicator was a 1 kHz tone that was on for 500 ms and off for 500 ms.

Airspace and Scenarios

We used generic, en route airspace previously designed to be realistic, yet relatively easy to learn [2]. We used the traffic scenarios from the earlier simulation [1] to generate scenarios for our practice and test scenarios. Each 60-minute test scenario consisted of high traffic volume with 94 total aircraft (54 arrivals, 11 departures, and 29 overflights). During each test scenario, the participants in the simulators flew 5 of the 94 aircraft in the scenario and the pseudopilots handled the rest. The participant in the B-747 flew in three flight segments (one overflight, one arrival, and one departure) as the pilot not flying (PNF) and was responsible for handing communications with the controller. The participant in the C-421 flew in two flight segments (one arrival and one departure) because of the slower speed of the C-421 aircraft through the sector. This participant served as the pilot flying (PF) and handled all communications.

Experimental Design

We used a two-factor design with two levels of the communications system (analog and digital) and two levels of the environmental condition (routine and weather). We manipulated both factors within subjects so that each participant experienced all four combinations of communications system and environmental test conditions. We counterbalanced the order of the radio conditions across participants to reduce any order effects. However, the participants always worked with a communications system under the routine condition before working with it under the weather condition. We also counterbalanced the order of the flight segments across participants.

Procedure

Two participants, one in the B-747 and one in the C-421, completed the practice and test scenarios over a one and one-half day period. On the first day, the researchers briefed the participants about the purpose of the study and the experimental procedures. We demonstrated the digital system communications delay and system features and then had the participants complete a short practice session.

During the test scenarios, each participant assumed responsibility of the designated aircraft as the aircraft entered the sector and completed all necessary procedures until it exited the airspace. Human Factors Specialists (HFSs) observed the participants in each simulator and recorded the time and type of communications attempted, any blocks or overrides, and other observations.

For the weather scenarios, we instructed the participants to assume high severity when the weather segments started and to make a request for a new clearance as soon as possible. The objective of these weather events was to maximize the demand on the communications channel. The PF in the B-747 simulator (a member of the experiment team) prompted the participant to request a deviation around a thunderstorm at specified points in the scenario. The simultaneous onset of turbulence in the C-421 signaled that participant to request a deviation. We provided the pseudopilots with scripted prompts to indicate when the weather events would occur and what communications needed to be made from the designated aircraft. We instructed the pseudopilots to call in their scripted reports of weather or requests for deviations as indicated by the time on their script sheets.

To minimize variability in controller performance, one supervisory en route ATCS from a field facility acted as the controller throughout the practice and test scenarios for each pair of

---

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

participants. Two controllers assisted during the course of the simulation.

During each scenario, the participants rated their overall workload after each flight segment. After each test scenario, the participants completed a questionnaire to rate their workload using the multidimensional NASA-TLX rating scale [3], performance, situation awareness, the extent to which the communications system affected various tasks, and the operational acceptability of the system using a modified version of the Controller Acceptability Rating Scale [4]. After the final test scenario, the participants completed an exit questionnaire to directly compare the radio systems and provide an overall assessment of the scenarios, system realism, and the VDL3 system features. They also discussed these and other issues in a final debriefing session with the HFS observers.

Objective Measures

During the simulation runs, we collected multiple objective measures to assess the effects of the communications systems and the weather. For the communications data, we analyzed the time and duration of each PTT action for the participants, controller, and pseudopilots and calculated the number of overlapping transmissions and the number of successful and unsuccessful transmissions in each scenario. Overlapping transmissions occurred when at least one other user keyed the microphone while another user was already keying. In the analog condition, any transmission attempt made while another is already keying would result in two (or more) unsuccessful transmissions. In the digital condition, a transmission attempt made by one pilot while another pilot or controller is already keying, would result in a block of that pilot's attempt, causing one unsuccessful transmission. If a pilot is already on the channel and the controller keys the microphone, this would result in an override, and one unsuccessful transmission. Controller overrides were categorized as either unintentional or intentional. We defined unintentional overrides as those that occurred within 360 ms from the start of a pilot PTT, in which it would not have been possible for the controller to hear the pilot begin to speak. We categorized intentional overrides as those that occurred more than 360 ms after the pilot keyed.

We also measured the total amount of time that the channel was occupied during the 60-minute scenarios. In the analog condition, the channel was determined to be occupied as long as at least one user was keying the microphone. To calculate channel occupancy in the digital condition, we summed the durations of each PTT from key press to key release, but excluded any transmissions or portions of transmissions that were blocked or overridden.

Data Analyses

We evaluated the communications measures on a per-scenario basis and used an alpha level of .1 as our criterion for determining statistical significance for these data because of the relatively small sample size (7 participant groups). We evaluated the subjective measures by individual and, therefore, set the alpha level to .05 because of the higher sample size.

We used both multivariate and univariate statistics to analyze the data. We used multivariate analyses when we determined that two or more variables made up measures of a higher order metric (e.g., workload). If the multivariate results were significant, we conducted univariate tests. We present only the univariate results for easier interpretability. We used proportions for many of our communications measures. Since proportions are often not distributed normally, we used an arcsin transformation to normalize those data. Analyses of both the proportions and the transformed data yielded equivalent results in terms of significance, so we present the proportions for easier interpretability.

Results

Communications

We compared the transmission rate and the rate of overlapping and unsuccessful transmissions between the radio condition, with and without weather. We found no statistically significant differences in the total number of transmissions, the number of controller transmissions, or the number of pilot transmissions made between the analog and digital radio conditions. On average, there were about 650 transmissions per 60-minute scenario, and the pilots made about 60% of them (see Figure 1). However, the total number of transmissions was significantly higher during the weather scenarios [$F(1, 5) = 55.64$, $p = .001$], as were pilot [$F(1,5) = 147.78$ $p < .001$] and controller transmissions [$F(1,5) = 9.14$, $p = .029$].

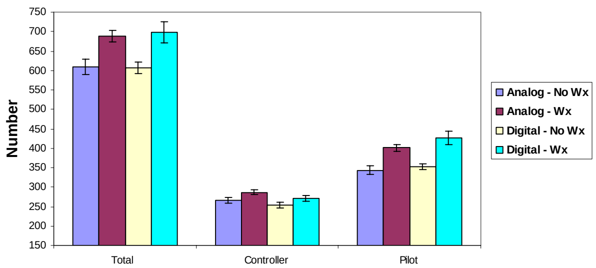

Figure 1. Transmission Rate

The amount of time that the channel was occupied also did not differ significantly between the analog and digital conditions (see Figure 2), but was significantly higher in the weather conditions [$F(1,5) = 19.40$, $p < .007$]. Overall, the radios were in use about 50-55% of the total scenario time.

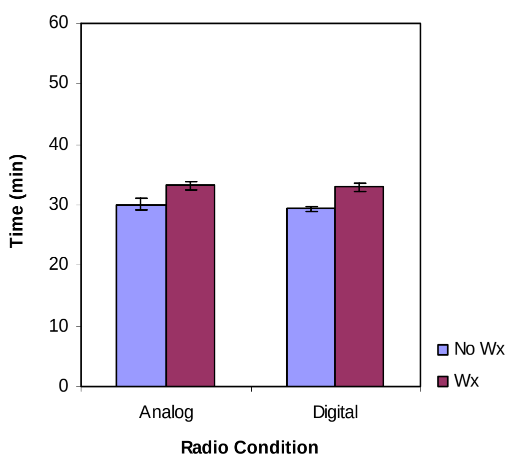

Figure 2. Channel Occupancy

The rate of overlapping transmissions did not differ between radio conditions (see Figure 3). About 18% of transmissions overlapped in the analog condition, and about 17% overlapped in the digital condition, suggesting that the way users attempted to access the channel did not differ based on which radio was used. However, the rate of overlapping transmissions was significantly higher when weather was involved [$F(1,5) = 104.69$, $p < .001$].

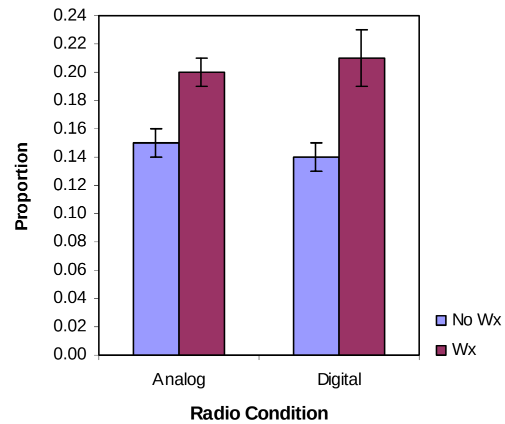

Figure 3. Overlapping Transmissions

As expected, there were significantly more unsuccessful transmissions in the analog condition than digital [$F(1,5) = 129.97$, $p < .001$]. Approximately 18% of the total number of transmissions made were unsuccessful in the analog condition compared to about 10% in the digital condition (see Figure 4). Significantly more unsuccessful transmissions were also made when weather was involved [$F(1,5) = 77.08$, $p < .001$].

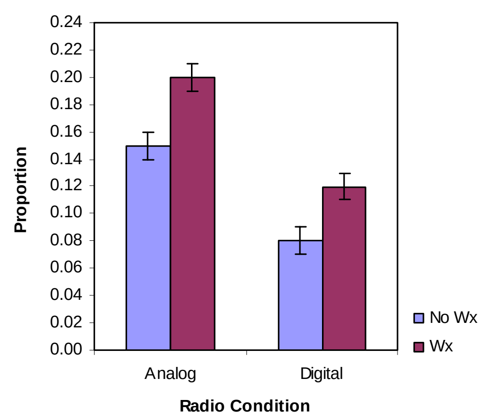

Figure 4. Unsuccessful Transmissions

We also examined the rate of unsuccessful pilot transmissions separately because only pilot transmissions can be unsuccessful with the digital system (see Figure 5). The proportion of unsuccessful pilot transmissions was significantly higher in the analog condition [$F(1,5) = 7.87$, $p = .038$] and significantly higher when there was weather involved [$F(1,5) = 85.92$, $p < .001$].

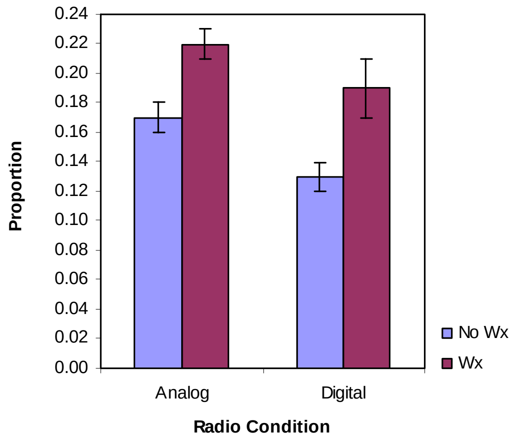

Figure 5. Unsuccessful Pilot Transmissions

We analyzed the characteristics of overlapping transmissions in detail to determine when pilots were blocked, stepped on, or overridden. Of the 237 controller overrides that occurred using the digital system, 65% were categorized as unintentional because they occurred within the first 360 ms of a pilot key press. Of those that were intentional, 88% were made within the subsequent 360 ms, or 720 ms from the start of the pilot key press. If we assume that the pilot did not begin to speak immediately upon keying the microphone, some of the overrides that we categorized as "intentional" may have actually been initiated before the controller heard the pilot speak.

We also looked at the comparable measures in the analog conditions, in which the controller stepped on a pilot transmission. We found that 14% of controller step-ons occurred during the first 70 ms of a pilot key press, the length of the voice throughput delay, during which time it would not have been possible for the controller to hear the pilot begin to speak. Overall, 32% of the 327 controller step-ons occurred within the first 360 ms of the pilot key press, relatively early into the pilot transmission. Unlike the digital condition, about half of the step-ons occurred more than 2 seconds after a pilot keyed the microphone. Of these, 64% occurred within the last 360 ms of a pilot transmission. This result is not surprising because it is unlikely that controllers would deliberately step on another transmission when that attempt would not allow them access to the channel. It is possible in these situations that the controller was attempting to access the channel after the pilot had finished talking but had not yet released the microphone key.

We also analyzed pilot attempts to access an occupied channel in both radio conditions. In the digital condition, this results in a block. In the analog condition, the attempt would result in a step-on. Though the uplink and downlink delays do not apply between pilot-pilot communications, we were interested in determining at what point into an existing transmission another channel access attempt was made. We found that 72% of pilot attempts to access an already occupied channel occurred within the first 720 ms of an existing transmission.

Workload

We averaged the workload ratings for each participant across flight segments to obtain a single measure of overall workload for each test condition. We also obtained workload ratings for several dimensions (mental, physical and temporal workload, effort, and frustration) on the NASA-TLX questionnaire. We analyzed these data using a multivariate analysis of variance (MANOVA) and found no statistically significant difference in the ratings across test conditions. Overall, the workload ratings tended to be low, averaging about 2 on a 10-point scale, except for the "effort" measure which was somewhat higher (mean = 4.1), indicating a more moderate workload level.

We also evaluated participant ratings for communications tasks to assess the extent to which the different systems used interfered with those activities. Using 10-point scales (1 = not at all; 10 = a great deal), the participants rated the effect of each system on their piloting strategy, ability to make timely transmissions, receipt of critical air traffic services, receipt of optional air traffic services, and use of standard communication procedures. In addition, the participants rated how much each system affected their speech and the number of communication problems or mistakes. Almost all of the ratings had averages less than three, indicating that there was only a small negative effect attributed to either system. A MANOVA revealed no significant difference in ratings across test conditions.

System Acceptability

The participant acceptability ratings for each system were high, averaging between 8.7 and 9.6 on the 10-point scale for the analog and digital system, respectively. Both systems were considered highly acceptable and had minimal deficiencies, but the digital system ratings were significantly higher than the analog system ratings [$F(1,12) = 21.13$, $p = .001$]. The difference in the ratings between the environmental conditions was not significant.

System Comparison

At the end of the study, the participants compared the two radio systems on 10 communications functions: completing routine pilot-initiated radio calls, completing time-critical communications, responding to air traffic controller calls, determining when the channel was busy, determining when the channel was available, being confident the controller received the message, detecting communications problems, hearing complete messages, receiving timely responses, and accomplishing all communications tasks. The participants indicated the extent to which they found one system “much better,” “somewhat better,” or “no different” than the other in supporting their ability to perform a specific task. All but two of the 140 individual ratings indicated that the participants perceived the digital system as equivalent to or better than analog. Sixty-six percent of the responses indicated that the digital system supported these functions “somewhat better” or “much better” than the analog system.

Digital System Features

The participants rated the usefulness of the antiblocking, controller override, and TSI very highly, with average responses of 9.5, 9.2, and 9.3, respectively.

The extent to which the antiblocking and controller override features negatively affected communications was fairly low, with average ratings of 3.5 and 2.4, respectively, on 10-point scales in which 10 indicated an extremely negative effect. However, these ratings were variable. While nine of the participants indicated the negative effect of antiblocking was low (ratings of three or less), four participants indicated the negative effect was moderate (ratings between five and seven). The remaining participant rated the effect a 10, but we believe this may have been a misinterpretation of the rating scale because the same participant also rated the usefulness of the feature a 10.

Ten participants rated the negative effect of controller override very low (ratings of one or two), while the remaining four participants rated its negative effect as low to moderate (ratings of four and five).

The participant ratings of the acceptability of the volume, pitch, and on-off cycle of the TSI varied widely. Though responses averaged about 7 on a 10-point scale (1=lowest, 10=highest), they ranged from 3 - 10 suggesting some improvements may be useful. The tone was more difficult to hear in the C-421 because the engine noise (about 76 dBA) in that simulator was louder than in the B-747 (less than 60 dBA). In addition to the ratings and comments received, we were concerned that a single frequency TSI would be difficult for people with hearing impairments to detect, particularly amid loud ambient noise.

Transmit Status Indicator Evaluation

In a subsequent analysis, members of the NEXCOM Human Factors Working Group (HFWG) evaluated three TSI implementations to determine which was most detectable in ambient cockpit noise, but did not mask the intelligibility of incoming voice messages. One TSI was the originally specified 1-kz tone with a 500 ms on-off cycle. The other two were multifrequency tones consisting of 500 Hz, 1 kHz, 2 kHz, and 4 kHz.

Multifrequency tones are more likely to be detected by users who have hearing deficiencies in part of the auditory spectrum. One multifrequency TSI was presented with a 200 ms on, 300 ms off cycle. The other multifrequency TSI was presented with a 300 ms on, 700 ms off cycle.

Each TSI was presented with 85 dBA jet engine or propeller aircraft background noise. We presented the stimuli in an audio booth through a speaker to simulate a "worst case" condition for detecting the TSI in ambient noise. Six listeners participated in a two-part evaluation to assess TSI detectability and the extent to which the TSI affected the intelligibility of underlying messages[2].

In the detectability evaluation, the TSIs were presented in ascending and descending series in 5 dBA increments from 60 - 85 dBA in random order. Each TSI was presented for 1.5 s at each increment. The presentation times of TSIs within each series were randomized between 3 and 8 s to reduce anticipation. The listeners pressed a button each time they heard a TSI. We calculated the detectability thresholds for each participant and determined the average threshold per condition (see Table 1). Lower threshold values indicate better detectability.

Table 1. Detectability Thresholds

Back-ground	500/500 ms on/off mean (sd)	200/300 ms on/off mean (sd)	300/700 ms on/off mean (sd)
Prop	67.9 (3.32)	60.6 (1.39)	60.6 (1.36)
Jet	72.5 (1.73)	69.4 (1.02)	69.0 (1.16)
Totals	70.2 (2.14)	65.0 (1.05)	64.8 (1.04)

The three TSI thresholds differed significantly from one another [$F(1,4) = 33.81$, $p < .005$]. Overall, the TSIs were more easily detected in propeller background noise than jet noise [$F(1,4) = 1383.87$, $p < .001$], but the trends across the three TSIs did not differ significantly between the different backgrounds.

With propeller background noise, nearly all of the responses were correct for the multifrequency TSIs. Four of the five listeners heard all the tones at each dBA level. The other listener heard 93% of these tones. There were no false alarms reported. With jet background noise, there were three 1 kHz TSI series (out of 64) in which none of the TSIs were heard. Nine false alarms were reported, seven of which occurred for the 1 kHz TSI, and two of which occurred for the multifrequency, 200/300 ms on-off cycle TSI. Overall, the 1 kHz TSI with 500 ms on/off cycle was the most difficult to detect.

The second phase of the evaluation involved determining the intelligibility of ATC messages in the presence of the TSI. We used only the more detectable multifrequency TSIs in this phase of the evaluation and presented them over recorded ATC voice messages (e.g., "American 128, turn left heading 190 at BATTS intersection.") in 80 dBA ambient jet noise. One half of the voice messages were spoken by a male, the other half by a female. Two TSI amplitude levels were evaluated. One was at the same amplitude as the voice (0 dBA relative), and the other was 5 dBA below (-5 dBA relative) the voice amplitude. Before the session started, each listener heard 6 ATC messages without TSIs and adjusted the volume of the messages to a desired level. No adjustments were made once the evaluation started. The median voice setting was 85 dBA and ranged from 82 to 93.

After a message was played, the stimulus tape was stopped and the listener filled in three blanks on an answer sheet to indicate what was heard (e.g., "American _____, turn left heading ____ at ______ intersection.").

Presentation orders of TSIs, the relative amplitude levels (0 or -5 dBA), and sex of voice were randomized. Overall, an average of 12 correct responses (sd = 1.8) were made out of 15 (see Table 2).

Table 2. Intelligibility Scores

	200/300 ms on/off		300/700 ms on/off	
Relative TSI amplitude	**Female mean (sd)**	**Male mean (sd)**	**Female mean (sd)**	**Male mean (sd)**
0 dBA	11.0 (2.19)	11.3 (2.06)	12.2 (1.17)	11.2 (1.94)
- 5 dBA	12.2 (2.32)	13.0 (1.1)	12.5 (.84)	13.3 (1.21)

Intelligibility was better for the - 5 dBA TSI (mean = 12.8, sd = .73) than for the 0 dBA TSI (mean = 11.4, sd = 1.26) at the .1 level [$F(1,5)=5.16$, $p = .072$]. We used an alpha level of

[2] Data from five listeners were used in the detectability evaluation because data were corrupted for the sixth.

.1 as our criterion for determining statistical significance for these data because of the relatively small sample size (n = 6).

The listeners also rated whether the tone heard during each message was (1) much too quiet, (2) a little too quiet, (3) just right, (4) a little too loud, (5) much too loud. There was no difference in the loudness ratings between the two TSIs at the same amplitude. However, the loudness ratings for the two TSI amplitudes differed significantly [$F(1,5) = 11.421$, $p = .02$.], with ratings for the 0 dBA relative tone tending more towards "a little too loud" than those for the -5 dBA tone (see Table 3).

Table 3. TSI Loudness Ratings

	200/300 ms on/off		300/700 ms on/off	
Relative TSI amplitude	**Female mean (sd)**	**Male mean (sd)**	**Female mean (sd)**	**Male mean (sd)**
0 dBA	3.3 (.22)	3.4 (.47)	3.3 (.35)	3.4 (.36)
- 5 dBA	3.1 (.47)	2.9 (.37)	3.1 (.21)	3.0 (.46)

We also found a significant interaction of TSI amplitude x sex of voice [$F(1,5) = .942$, $p = .037$]. The difference in the loudness ratings between the 0 and -5 dBA amplitudes was greater when messages were presented by the male speaker.

We then examined other factors that could have affected the results. First, we evaluated the effect of the test item types to determine whether less familiar items, such as specific location names (e.g., "Savannah") contributed more to errors than did more common, predictable, items such as call signs and altitudes. We eliminated the less familiar items from the analysis, but found no significant differences in the proportions correct relative to the full data set.

Second, we evaluated whether practice effects contributed to the results. We compared the first five messages in a sequence with the last five messages and found that performance was significantly better on the last five messages, suggesting that participants had more difficulty when first starting the task [$F(1,5) = 17.093$, $p = .009$].

Finally, we found that more errors were made in the third test item position (60 out of 140 total errors) compared to those made in the first and second test item positions (40 in each), suggesting that forgetting may have played a role. This finding is in agreement with some of the listeners' comments that items were often understood but not remembered when they filled in the blanks.

Overall, our findings for the TSI evaluation indicated that the single frequency TSI was harder to detect than the multifrequency TSIs. A multifrequency TSI with on/off cycles between 200/300 and 300/700 ms and approximately 5 dBA lower in amplitude than the voice would be an effective alternative.

Discussion

The results showed that the digital radio system allowed more successful transmissions to be made than the analog system, even though users tended to access the channel similarly with both radio systems. Communications increased in the weather conditions, but did so similarly for both radio systems. The participants were highly positive in their evaluations of VDL3, rating the operational acceptability of the digital system significantly higher than that of the analog system regardless of the weather. The digital system was almost always rated as being equal to or better than the analog system for completing communications tasks. The antiblocking, controller override, and TSI features were also rated highly useful, though a separate evaluation indicated that a multifrequency TSI would be more easily detected than the proposed 1 kHz tone and would allow for good intelligibility of the underlying voice messages. Overall, the results indicate that the VDL3 system is an acceptable communications system for pilots.

References

[1] Sollenberger, Randy, D. Michael McAnulty, Karol Kerns, 2003, The effect of voice communications latency in high density, communications-intensive airspace: Final report, DOT/FAA/CT/TN03/04, Atlantic City International Airport, NJ, William J. Hughes Technical Center.

[2] Guttman, Joseph A., Earl S. Stein, 1997, En route generic airspace evaluation, DOT/FAA/CT-TN97/7, Atlantic City International Airport, NJ, William J. Hughes Technical Center.

[3] Hart, Sandra G., Lowell E. Staveland, 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.

[4] Lee, Katherine K., Karol Kerns, Randall Bone, 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.