Evaluating Voice and Data Link
Air Traffic Control Communications for General Aviation

Randal J. DeMik and Bruce W. Welsh
Indiana State University

ABSTRACT

This study examined the effects of data link and voice air traffic control commands on pilot recall and execution. Instrument-rated pilots ($N = 26$) were tested on both a series of data link command tasks and a series of voice command tasks. The researchers predicted that participants would have a significant difference in overall errors in pilot recall and execution in the data link ATC command condition as compared to the voice ATC command condition. Also predicted was that there would be a greater gap in errors in pilot recall and execution at the ATC command blocks that contained higher parameters of ATC instructions. Our results indicate that pilots had significantly fewer errors in recall and execution in the data link condition compared to the voice condition at the moderate and high load ATC command levels. There was little or no difference in errors in pilot recall and execution in the data link condition compared to the voice condition at the low load ATC command levels.

INTRODUCTION

The world-wide aviation community is interested in implementing options to the traditional interaction via voice exchanges between air traffic controllers and pilots. One of these options is Controller to Pilot Data Link Communication (CPDLC). CPDLC uses a Very High Frequency (VHF) or satcom link to route text messages that are displayed on Flight Management System (FMS) or Aircraft Communication Addressing and Reporting System (ACARS) screens in the cockpit (Ambrose, 2004). According to Kerns (1991), the perceived benefits of data link include an improved clarity and efficiency of communication, reduced number of misunderstood communications, expanded airspace capacity, freeing up of frequencies used in voice communications, and reduced pilot and controller workload.

Today, the bulk of exchange between controllers and pilots is carried out by means of voice communication. Flight operations in the National Airspace System (NAS) depend on the timely and accurate exchange of information between Air Traffic Control (ATC) and pilots in the cockpit (McGann, Morrow, Rodvold, & Mackintosh, 1998). According to Helleberg and Wickens (2003), "…data link is one of the new technologies designed to replace or alter more traditional information exchanges between the

pilot and ATC" (p. 1). Wickens, Mavor, Parasuraman, and Mcgee (1998) reported on the challenges facing the NAS and determined the need to upgrade the system within the context of the next generation air traffic control system (NGATS). NGATS and *free flight* will primarily use orbiting satellites, on board automation, and data link communications.

LITERATURE REVIEW

In a study by Olson (1996) to determine the services that general aviation pilots desired via data link, he found a high preference for data link use in Pilot Reports (PIREPS), Notice(s) to Airmen (NOTAMS), Automated Terminal Information Service (ATIS), and instrument flight rules (IFR) operations. In a NASA study (Lee et al., 2003) that focused on ATC controllers' views of data link, researchers found that "controllers had a high preference for transfer of communication through data link as a workload saving mechanism" (p. 1). Despite these early studies that highlighted both controller and pilot preference for text-based technologies, current use of data link in the NAS is limited to pre-departure clearances and oceanic clearances between ATC and airline crews via a third party delivery system (Ambrose, 2004).

According to Ambrose, the FAA began direct data link trials in a Miami Center test

program in 2003, with a complete rollout for all national high-altitude control centers planned for 2006. Due to lack of funding, however, this program was put on hold in late 2003. Despite the setback in the U.S. CPDLC project, the Europeans are continuing with their rollout of the system. Currently, data link is being used at the Europe's Maastricht Upper Area Control Center, which handled over 4,000 CPDLC flights in a recent 12 month period (Hughes, 2005). The Maastricht project includes support and participation from U.S. airlines and U.S. avionics companies. Hughes summarizes that the complete ATC deployment, as well as the mandate of data link avionics in aircraft, is planned for the entire European continent by 2009.

According to Kerns (1991), past experience with the advent of innovative automation applications to the flight deck have indicated that changes in technology and machine responsibility will alter workload demands on pilots, and that additional research will be needed to determine human performance, technology design, human-computer interaction, and future training requirements. Wickens et al. (1998) furthers this notion that while these innovative technologies may improve efficiency, they create additional concerns about human performance integration with automation systems design. Wickens et al. conclude that the choice of what to automate should be guided by research in human-centered automation that focuses on the need to compensate for human vulnerabilities.

According to earlier research, most of the results of previous data link studies focused on airline crews and airline operations (Billings & Cheaney, 1981; Lee, 1989; McGann, Morrow, Rodvold, & Mackintosh, 1998). With the advent of very light jets (VLJs) and the use of multi-function displays with data link capabilities in current light piston Technically Advanced Aircraft (TAA), general aviation now has the ability to use data link in their operations. Few studies have compared the two delivery methods of voice and text in today's general aviation system.

One study (Risser, Scerbo, Baldwin, & McNamara, 2006) used non-pilot graduate students to manipulate a panel using a computer screen and mouse controls in response to simulated speech and text ATC commands while measuring response time and accuracy. Another study (Helleberg, Wickens, & Goh, 2003) used fifteen instrument rated pilots in a simulator with a visual display focusing on heads-down time results while scanning for traffic using three different data link display conditions. A third study (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003) used twelve instrument rated pilots in a flight simulator that incorporated data link and cockpit display of traffic information while primarily reporting results on visual scanning.

PURPOSE

Few studies have compared the two methods of ATC command delivery (voice vs. data link) in a modern general aviation cockpit environment. Empirical studies of general aviation pilot performance with data link are limited, especially with respect to future requirements. Therefore, the purpose of this research study was to evaluate and measure accuracy of pilot recall and execution regarding use of text (data link) or voice as an ATC communications interaction in general aviation.

The researchers first predicted that there would be no main effect or interaction effect for order of trial (voice first or data link first). Then we predicted that participants in this study would have a significant difference in overall errors in pilot recall and execution in the data link ATC command condition compared to the voice ATC command condition. Finally we predicted that there would be a greater gap in errors at the ATC command blocks that contained a higher number of parameters of ATC instructions.

METHOD

Participants
Participants included 26 instrument-rated pilot volunteers from Indiana State University who all held a current FAA medical. Those who volunteered, and did not hold at least a current third class FAA medical were excluded, as were potential participants that were not instrument-rated. One condition of this experiment's design

was that participants have no auditory, visual, or other impairments that may affect the results of this study. This condition was controlled for by requiring all participants to hold at least a current third class FAA medical. The pilots' total flight hours ranged from 160 to 1,220 hours ($M = 350.0$, $SD = 244.8$).

Measures

The Frasca 142 flight training device (FTD) located in the Indiana State University Flight Simulator Lab was used to conduct these trials. The FTD was preset to conform to the flight characteristics of a Cessna 172 Skyhawk RG. The FTD consisted of a primary flight instrument panel arranged in the standard 'T' configuration. The radio control panel included two communication radio controls, two navigation radio controls, and a transponder control head. The pilots flew the FTD manually (no autopilot) using traditional aircraft controls (control yoke, rudder pedals, and a single throttle control). A Dell 15-inch color monitor set at a screen resolution of 1024 X 768 was mounted in the FTD in the pilot's normal field of vision. The monitor displayed a data link control panel measuring approximately 4 inches by 5 inches. An aural ACARS tone was used to alert participants of an incoming text message.

Procedures

At an initial briefing, participants gave informed consent for this Institutional Review Board (IRB) approved study. Each participant flew two cross-country flights (each approximately 15 minutes long) under instrument flight rules (IFR) in simulated instrument meteorological conditions (IMC). Each trial consisted of two flights: a flight under the ATC voice command condition; and another flight under the ATC text (data link) command condition. A counter-balancing technique was used to control for the carry over effects of practice effect or positive carry over treatment effect. In this study, half the participants were given the voice command task flight first and the text (data link) command task flight second. The other half of the participants were given the text (data link) task flight first and the voice command task flight second.

The pilots began each simulated flight at a preset altitude, airspeed, and heading. The scenario began as a flight that was airborne after having departed a local non-tower airport and requesting an IFR clearance in the air. Once the participant was comfortably established in stable cruise flight, the FTD operator initiated ATC commands regarding initial clearances (heading, altitude, airspeed, IFR clearance, squawk, and altimeter setting). In the text condition, ATC commands were always preceded by an aural alerting tone as an indicator of an incoming data link message. In the voice condition, pilots were briefed to either verbally read back or acknowledge all ATC commands. In the text condition, pilots were required to press an acknowledge button near the data link display to confirm receipt of a message and willingness to comply with ATC commands. In the voice condition, the FTD operator acted as a pseudo-controller and read the ATC scripted commands to the pilot. The participants were briefed that they could ask to have controller commands repeated to them. The FTD operator, acting as the pseudo-controller, would also correct participant's errors during acknowledgements and read backs.

Each flight included nine ATC command blocks. Three ATC command blocks had a load of four parameters or more in length and were considered high workload (e.g., Frasca 142 is cleared to the STL airport except fly heading 280 to intercept the VLA 250 radial via the VLA 4 arrival, climb and maintain 5,000 ft., squawk code 2312). Another three ATC command blocks had a load of three parameters in length and were considered moderate workload (e.g., Frasca 142 turn right heading 360 descend and maintain 4,500 ft., and maintain 110 knots.). The other three ATC command blocks had a load of one or two parameters in length and were considered low workload (e.g., Frasca 142 contact St. Louis Approach on 126.5).

The main factors of interest in this study were the number of errors made in pilot recall and execution across the ATC command condition (voice or text) given the level of command workload (high, moderate, or low) and order of flight presentation (voice or text first). An observer counted the number of errors made by each pilot in ATC command recall and execution. For example, if the pilot was given a command to descend to an altitude and the pilot

either did not initiate the descent or set the wrong altitude (misread, misheard or did not execute properly) the observer recorded an error.

The characteristics of this research study are consistent with a within-subjects design where each of the participants provided data from two trials (flights), each under a different condition (voice or text ATC commands). The order of trials was altered using a counterbalancing technique which also provided a between-subjects factor. All data for this study was entered into SPSS 14.0 for analysis. A *3* (levels of ATC commands) X *2* (conditions of voice or text) X 2 (order of trials) mixed model Analysis of Variance (ANOVA) design technique was used to provide the initial analysis for this study. The dependent variable was the number of errors in pilot recall and execution as recorded by the observer. The independent variables were condition of ATC commands (voice or text), loads in parameters of commands (high, moderate, or low), and a between-subjects factor of order of trial (voice first or data link first).

The Type I error for this investigation was set at .05 (α = .05). Minium, Clarke, and Coladarci (1999) suggest that the Type I error of .05 is the most commonly used for this type of research. While this alpha gives a higher probability of Type I error than an alpha of .01, the results of this data link study are informational only. In this case, the risk involved in a Type I error is small and reducing the Type I error to .01 was not necessary.

RESULTS

A *3* X *2* X 2 mixed design ANOVA was used to first determine the between-subjects effects of the order of trial (voice first or data link first) on errors in pilot recall across the within-subjects effects of ATC command condition (text or voice) and ATC command load (high, moderate, or low). No significant main effects or interactions for order were found. The ATC Command condition X Order interaction (F (1, 24) = .937, p > .05) and the ATC Command condition X Load X Order interaction (F (2, 48) = .385, p > .05) were all not significant. The analysis fails to reject the first null hypothesis. Data link and voice command condition errors in pilot recall and execution were not significantly influenced by order of trial.

A *3* X *2* X 2 ANOVA was then used to determine the effect of the ATC command load (high, moderate, or low) and command condition (voice or text) on errors in pilot recall and execution. The descriptive statistics for the means of each condition are consistent with the researcher's prediction that participant errors in recall and execution at the high load ATC command blocks that contained four or more parameters of instructions would be less in the text (data link) condition than in the voice condition. The average overall number of errors made in pilot recall and execution while flying the flight training device under the data link command condition (M = 4.2) were less than the average number of errors while flying under the voice command condition (M = 6.8). There is little or no difference in errors in pilot recall and execution across voice or text at the lower parameters of commands (see Table 1).

The errors in pilot recall and execution were then analyzed within the *3* X *2* X 2 ANOVA across the command conditions of voice and text. A significant effect was found (F (1, 24) = 15.4, p < .05). The second null hypothesis was rejected. Overall number of errors in pilot recall and execution in the text (data link) condition are significantly less than in the voice ATC command condition. Further analysis then compared the pilot recall and execution errors in the conditions of voice and text to include the load levels of ATC commands (high, moderate, or low).

Table 1. *Descriptive Statistics for Mean Errors in Pilot Recall and Execution*

ATC Command Condition	Overall Errors	High Load Errors	Mod Load Errors	Low Load Errors
Voice (n = 26)	6.8	3.8	2.0	0.9
Data link (n = 26)	4.2	2.2	1.1	0.8

A significant effect was found (F (2, 50) = 9.1, p < .05). The third null hypothesis was also rejected. There is a significant difference in number of errors in pilot recall and execution with regard to load of ATC command parameters (high, moderate, or low) in either the text or voice condition (see Table 2).

Table 2. *ANOVA Table of Within-Subjects Effects*

Source	SS	DF	MS	F	p-value
Command	30.519	1	30.519	15.433	.001
Command X Order	1.853	1	1.853	.937	.343
Error (command)	47.462	24	1.978		
Load	126.27	1.7	73.307	32.213	.000
Load X Order	7.321	2	3.660	1.868	.166
Error (Load)	94.077	48	1.960		
Command X Load	15.500	2	7.750	5.890	.005
Com X Load X Order	1.013	2	.506	.385	.683

Upon examination of the data, it appeared that the high load ATC commands had fewer errors in pilot recall and execution in the data link condition. This effect appears to diminish at the low loads of ATC commands (see Figure 1).
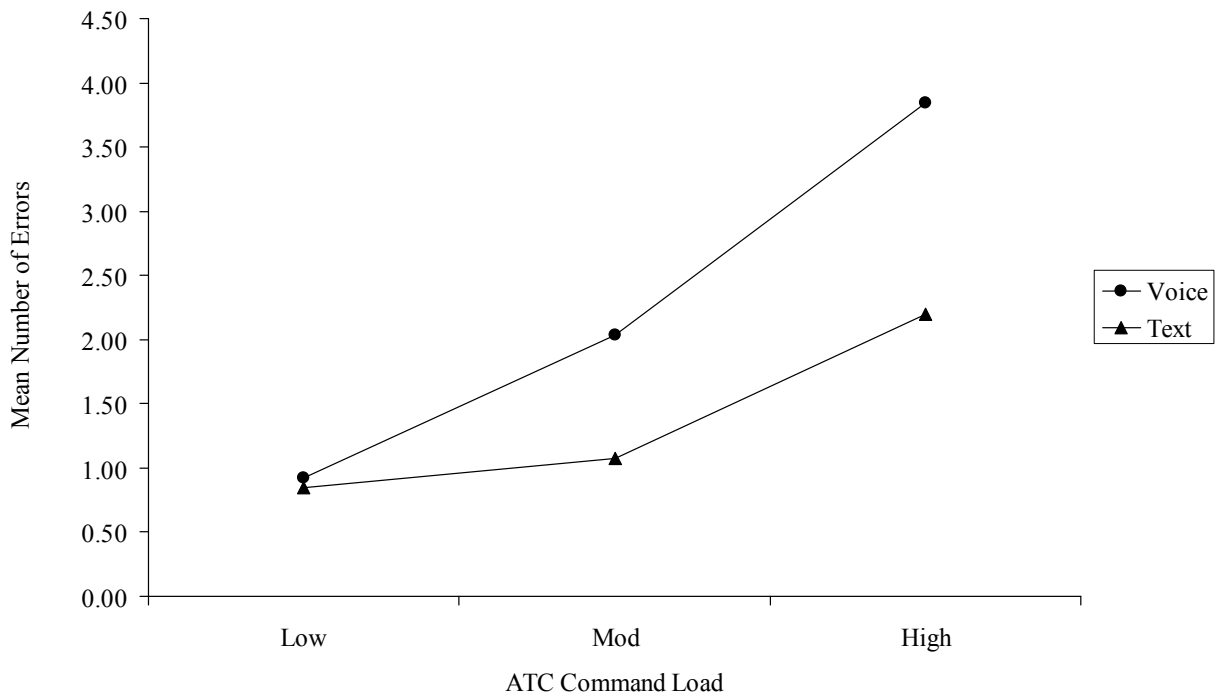

Figure 1. Errors in Pilot Recall and Execution across ATC Command Loads in Voice versus Text

As a follow-up, three protected dependent *t* tests were conducted to test that the mean errors in pilot recall and execution increased significantly with ATC command load intensity. This procedure was recommended by Cronk (2006) to substitute as a post-hoc analysis for this type of situation in repeated measures ANOVA. Because this analysis included three tests, and therefore inflated the Type I error rate, we used a significance level of .017 (.05/3) instead of .05. The follow-up protected *t* tests were calculated to compare the mean text (data

link) errors to the mean voice errors in the high, medium, and low load parameter ATC command levels (see Table 3). The mean difference between high load voice commands and high load text (data link) was 1.62 (SD = 2.53) errors. A significant decrease in errors from voice to text at the high load ATC command level was found (t (25) = 3.26, p < .017). The mean difference between moderate load voice commands and moderate load text (data link) was .96 (SD = 1.08) errors. A significant decrease in errors from voice to text at the moderate load ATC command level was found (t (25) = 4.56, p < .017). Finally, the mean difference between low load voice commands and low load text (data link) commands was .08 (SD = 1.23) errors. No significant difference from voice to text at the low load ATC command level was found (t (25) = .32, p > .017).

Table 3. *Paired Samples t Tests*

Source	M	SD	t	DF	Sig.
Voice High – Data High	1.62	2.53	3.25	25	.003
Voice Mod – Data Mod	.96	1.08	4.56	25	.000
Voice Low – Data Low	.08	1.23	.319	25	.753

In summary, a pair wise comparison of the three protected dependent *t* tests, with an adjusted alpha level of .017, revealed a statistically significant decrease in errors from voice to text at the high and moderate load ATC command level. However, no significant difference from voice to text at the low load ATC command level was found.

DISCUSSION

The primary purpose of this study was to investigate how different levels of datalink and voice commands affect errors in pilot recall and execution. Specifically, this study addressed the human reliability analysis of the data link automation issue in the general aviation environment. The researchers predicted that participants would have a significant difference in overall errors in pilot recall and execution in the data link ATC command condition as compared to the voice ATC command condition. Also, we predicted that there would be a greater gap in errors in pilot recall and execution at the ATC command blocks that contained higher parameters of ATC instructions. In this study the pilots had significantly fewer errors in pilot recall and execution in the data link condition compared to the voice condition at the moderate and high load ATC command levels. However, there is little or no difference in errors in pilot recall and execution in the data link condition compared to the voice condition at the low load ATC command levels.

The findings of this study are consistent with those of Risser et al. (2006), where results demonstrated an advantage in text commands with longer messages. Taken together, these findings suggest that data link may provide an increase in pilot performance with regard to high parameter communications from ATC such as: ATIS information; NOTAMS; PIREPS; initial IFR clearances; detailed route changes; and oceanic clearances. However, the evidence from this study and previous research in the field generally endorses the role of the dual modalities of voice and text over a preemption of one modality of communication.

Appropriate or likely uses of this research include assistance in the development of flight performance objectives for future general aviation aircraft systems. Other uses include direction in the incorporation of data link technologies in general aviation and Air Traffic Control. While this study focused on comparing a pure voice to a pure data link medium of communication, future research could examine the dual use of voice and data link for general aviation. Another limitation of this study was a focus on one aspect of data link with regard to errors in pilot recall and execution. Further research would be necessary to determine other implications of a change from voice to text displays with regards to heads-down time and the possible negative implications of this technology for the high-priority visual tasks regarding single-pilot operations that are predominant in general aviation.

In conclusion, accurate pilot recall and execution is likely to remain one of the critical aspects in general aviation that will be challenged by new technologies. As a consequence, it is important to gather empirical evidence that will drive future cockpit technologies and pilot training programs to improve safety in general aviation. Wickens et al. (1998) summarize that these new tools should continue to be evaluated with human-centered simulation and careful experimental design. The introduction of the new data link technology into the general aviation cockpit should proceed gradually, with a high degree of attention to training, differences and pilot requirements.

ACKNOWLEDGEMENTS

REFERENCES

Ambrose, K. (2004, March). Controller pilot datalink communications. *Business and Commercial Aviation, 94,* 56-58.

Billings, C. E., & Cheaney, E. S. (1981). *Information transfer problems in the aviation system* (NASA No. 1875). Moffett Field, CA: NASA Ames Research Center.

Cronk, B. C. (2006). *How to use SPSS: A step-by-step guide to analysis and interpretation* (4th ed.). Glendale, CA: Pyrczak Publishing.

Helleberg, J., & Wickens, C. D. (2003). Effects of data link modality on pilot attention and communication effectiveness. *The International Journal of Aviation Psychology, 13*(3), 189-210.

Helleberg, J., Wickens, C. D., & Goh, J. (2003). *Traffic and data link displays: Auditory? visual? or redundant? A visual scanning analysis*. Paper presented at the meeting of the 12th International Symposium on Aviation Psychology. Dayton, OH.

Hughes, D. (2005, March 14). Europe's CPDCL goal. *Aviation Week & Space Technology, 162,* 84.

Kerns, K. (1991). Data-Link communication between controllers and pilots: A review and synthesis of the simulation literature. *The International Journal of Aviation Psychology, 1*(3), 181-204.

Lee, A. T. (1989). *Display-based communications for advanced transport aircraft* (NASA No. 102187). Moffett Field, CA: NASA Ames Research Center.

Lee, P. U., D'Arcy, J. F., Mafera, P., Smith, N., Battiste, V., Johnson, W., et al. (2003). *Trajectory negotiation via data link: Evaluation of human-in-the-loop simulation* (NASA No. 262-4). Moffett Field, CA: NASA Ames Research Center.

McGann, A., Morrow, D., Rodvold, M., & Mackintosh, M. (1998). Mixed-media communication on the flight deck: A comparison of voice, data link, and mixed ATC environments. *The International Journal of Aviation Psychology, 8*(2), 137-156.

Minium, E. W., Clarke, R. C., & Coladarci, T. (1999). *Elements of statistical reasoning* (2nd ed.). New York: John Wiley & Sons, Inc.

Olson, R. R. (1996). General Aviation Survey Analysis. *National Technical Information Service, 96*(11), 1-29.

Risser, M. R., Scerbo, M. W., Baldwin, C. L., & McNamara, D. S. (2006). Interference timing and acknowledgement response with voice and datalink. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. San Francisco, CA.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors, 45*(3), 360.

Wickens, C. D., Mavor, A. S., Parasuraman, R., & Mcgee, J. P. (Eds.). (1998). *The future of air traffic control: Human operators and automation.* Washington, D.C.: National Academies Press.