NextGen Flight Deck Data Comm: Auxiliary Synthetic Speech – Phase I

Tracy Lennertz, Judith Bürki-Cohen, Andrea L. Sparko
USDOT/RITA/Volpe Center

Nickolas Macchiarella, Jason Kring, Mike Coman, Tom Haritos, Jeffry Alvarado
Embry-Riddle Aeronautical University

Data Comm—a digital, text-based controller-pilot communication system—is critical to many NextGen improvements. With Data Comm, communication becomes a visual task. Although Data Comm brings many advantages, interacting with a visual display may yield an increase in head-down time, particularly for single-pilot operations. This study examined the feasibility of supplementing the visual Data Comm display with an auxiliary synthetic speech presentation. Thirty-two pilots flew two experimental scenarios in a Cessna 172 Flight Training Device. In one scenario, ATC communication was with a text-only Data Comm display; in the other, the text Data Comm display was supplemented with a synthetic speech display annunciating each message (i.e., text+speech). In both scenarios, pilots heard traffic with similar call signs on the party line and received a conditional clearance; however in just one scenario (counter-balanced between communication conditions), pilots received a clearance that was countermanded by a live controller before it was displayed. Results indicated that relative to the text-only display, the text+speech display aided single-pilot performance by reducing head-down time; and it may have prevented participants from acting prematurely on the conditional clearance. Supplementing text Data Comm with speech did not introduce additional complications: participants were neither more likely to erroneously respond to similar call signs, nor to ignore a live ATC voice countermand. The results suggest that the text+speech display did not hinder single-pilot performance and offered some benefits compared to the text-only display.

INTRODUCTION

Traffic in the National Airspace System (NAS) is increasing, and consequently, the radio frequencies carrying voice communications between air traffic controllers (ATC) and pilots are becoming increasingly congested. The transmissions may be noisy or broken up by pilots “stepping on” each other’s communications, or pilots may be prevented from notifying ATC in the case of an emergency because of frequencies blocked by stuck microphones, with potentially serious consequences. The fast pace of live-voice communications may inhibit proper readback or requests for clarification of ATC instructions or pilot queries. In addition, some properties of speech make voice communication particularly difficult. For example, pilots must listen for their call sign to receive ATC instructions, amid instructions to other aircraft with similar sounding call signs (e.g., 345 vs. 354). Such similarity can result in communication errors (Grayson & Billings, 1981). Moreover, controllers do not always adhere to standard phraseology (Bürki-Cohen, 1996; Cardosi, 1993) and vary in both accent and speech rate. Communication errors are even more likely with long or complex voice instructions (Bürki-Cohen, 1996; Cardosi, 1993). Even when an instruction is correctly heard, a pilot may later forget the information, write it down incorrectly, or erroneously enter the data into the Flight Management System (FMS; Kerns, 1999). Each of these factors can contribute to inefficient or inaccurate communication (Kerns, 1999).

Data Comm, a digital, text-based data communication system between pilots and controllers may alleviate some of these problems inherent in voice communication. Data Comm, a key enabler for many operational improvements envisioned in the NextGen Air Transportation System (NextGen), is expected to help accommodate the anticipated increase in airspace capacity by increasing the efficiency and accuracy of controller-pilot communication. With Data Comm, written messages are exchanged directly between ATC and a specific flight crew. Consequently, a flight crew can no longer appropriate such a clearance intended for another aircraft with a similar call sign on the shared frequency. Difficulties understanding messages due to speech rate or accent become non-existent. Messages are pre-formatted, reducing workload and promoting the use of standard phraseology. Data Comm also alleviates the flight crew’s reliance on memory; messages are stored in a log and pilots can read or retrieve the message when needed. Compared to voice communication, Data Comm is associated with fewer memory errors, and this benefit is most pronounced with long instructions (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003).

Replacing today’s radio communications, however, may result in some unintended consequences. For example, Data Comm shifts communication from an auditory to a visual task. This may lead to a potentially unsafe increase in head-down time, as pilots must interact with a visual display to read and respond to a message. Moreover, the increased visual task load may reduce flight precision and lengthen the time between ATC communication and flight crew response.

To address these concerns, the National Research Council (NRC) recommended that Data Comm should “[e]mploy redundant voice synthesis of uplink messages as a design option, operated in parallel with visual (text and graphics) display of the message” (Wickens, Mavor, Parasuraman, & McGee, 1998, p. 251). Moreover, the FAA was mandated to “…address the problems and concerns raised by the National

Research Council…[and] respond to the recommendations" (Title 49 United States Code Section 44516). Supplementing visual Data Comm with synthetic speech may mitigate some of the potential risks associated with text-only Data Comm, but it must do so without introducing new risks. Indeed, past research has yielded mixed results (cf. Helleberg & Wickens, 2003; Lancaster & Casali, 2008; McCarley, Talleur, & Steelman-Allen, 2010). The current study sought to clarify some of the results of previous research and respond to the concerns raised by the NRC by examining the feasibility of supplementing a visual Data Comm display with synthetic-speech annunciations in the single-pilot environment.

In the current study, each participant flew two experimental scenarios that varied in their implementation of Data Comm: ATC messages were presented via a text-only display in one scenario and via an auxiliary synthetic-speech display *in addition to* the text display (text+speech) in the other scenario. Each experimental scenario included a conditional clearance ("AT ORMOND VOR CLIMB TO 3,000"). Pilots may forget to wait for the condition or forget to respond once the condition is filled, but the likelihood of these errors may be decreased with auxiliary synthetic speech. Pilots also experienced one countermanded clearance in either the text-only or text+speech condition; with Data Comm, it is possible for a live controller to countermand a message by voice before it is displayed on the flight deck, and we sought to determine whether or not pilots would ignore the countermand when the subsequent yet outdated Data Comm message was reinforced by synthetic speech (a potential harmful effect). In both experimental scenarios, aircraft with similar sounding call signs were also heard on the party line; pilots may be less likely to respond to similar call signs with text-only Data Comm, but the presence of synthetic speech may refocus pilots on sound.

METHOD

Participants

Thirty-two (28 male, 4 female) commercial certificated and instrument-rated pilots at Embry-Riddle Aeronautical University participated in exchange for $20/hour. Participants ranged between 19-28 years of age (*M* = 22.3) and had an average total flight time (excluding simulator time) of 554.1 hours (*SD* = 427.1, range = 138-1950) and an average total loggable Flight Training Device (FTD) time of 159.7 hours (*SD* = 177.7, range = 13-737). Participants were run individually and the entire experiment took about 2.5 hours.

Apparatus

Participants flew a single-engine Cessna 172S (Skyhawk) FAA qualified Level 6 FTD developed by Frasca (220° x 60° visual display). Participants used a touch-screen tablet, attached to their kneeboard, to view and respond to Data Comm messages (see Figure 1). Data Comm messages were first announced via a clearly audible "ding dong." A highly-intelligible (see AT&T Natural Voices™, 2002) 16kHz synthesized female voice, AT&T's "Natural Voices Crystal," was used for the auditory Data Comm display. All auditory instructions were presented binaurally over passive-noise-reducing headphones allowing participants to hear communications over the sound of the simulated airplane. Pilots could adjust the volume of the headphones.

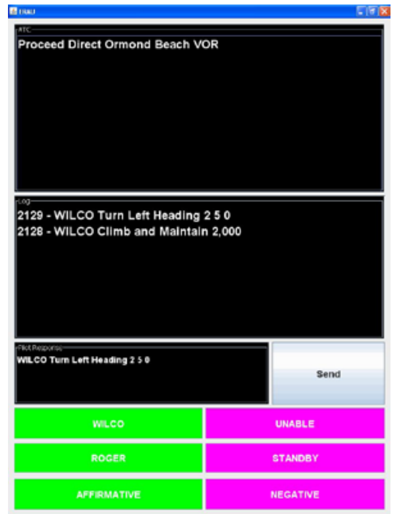

Figure 1. The touch-screen tablet.

Experimental Design

The experimental design included four independent variables: 1) Data Comm display modality (text only vs. text+speech), 2) call-sign similarity (similar vs. dissimilar), 3) type of Data Comm message (routine vs. conditional), and 4) modality of countermanded clearance (text only vs. text+speech). Data Comm display modality (text only vs. text+speech) and call sign similarity (similar vs. dissimilar) were examined together in a 2 by 2 completely within-subjects factorial design. The effect of Data Comm modality on conditional clearances was examined in a simple two-sample paired comparison (participants received a single conditional clearance amidst routine clearances in both the text-only and text+speech condition). The effect of Data Comm modality on the countermanded clearance was examined in a two-sample unpaired comparison (participants received one countermanded clearance in *either* the text-only or text+speech condition). A counterbalancing scheme was developed to avoid sequence effects.

Gaze-dwell time was recorded via two cameras on the instrument panel. Audio tapes recorded ATC interactions. Participant compliance with ATC instructions was determined via experimenter observation and objective flight precision measurement. Pilot response time to Data Comm messages was measured through 1) inputs to the touch-screen tablet; 2) inputs to the flight controls (where applicable); and 3) time to complete ATC instructions (where applicable). Flight precision data were collected through the FTD at a sampling rate of 30 Hz. Pilot opinion data (on workload, usability, helpfulness) were collected through surveys administered electronically after each experimental scenario and upon completion of the experiment.

Procedure

Each participant flew two 30-min (approximately) identical scenarios, in counterbalanced order. They departed and landed at Daytona Beach International Airport (KDAB) without leaving terminal radar approach control (TRACON). The scenarios were designed to impose moderate workload (i.e., they began in Visual Flight Rules [VFR] conditions and transitioned to Instrument Flight Rules [IFR] conditions). Traffic represented a high-activity day at KDAB. Data Comm was limited to Departure and Arrival ATC; all communications with the tower were via voice over radio. The amount of communication was designed to represent approximately 80% of the voice traffic occurring on the busiest days at KDAB. In both scenarios, two aircraft on the party line (i.e., 354 Echo Romeo and 345 Delta Bravo) had a call sign similar to the participants' ownship (i.e., 345 Echo Romeo).

While the actual flying was identical in both experimental scenarios, the presentation of Data Comm messages was varied. In one scenario, ATC instructions were issued via a Data Comm text display only. In the other scenario, instructions were issued via a Data Comm text display and annunciated by a synthetic voice. Each script contained mainly routine Data Comm messages sampled from the proposed RTCA SC-214/EUROCAE WG-78 message set, and one conditional clearance (e.g., "AT ORMOND VOR CLIMB TO 3,000"). Each scenario included 14 Data Comm messages: Six of the messages required the pilot to make a change to the flight controls (Key Events, e.g., "Turn Left Heading 310"); the remaining eight messages did not require the pilot to make a change (Stability Events, e.g., "At Dongs expect radar vectors for ILS 7 Left"). Participants experienced one Data Comm instruction that was countermanded by a recording of a live controller. One half (N = 16) of the participants experienced the live controller countermand in the text-only Data Comm display condition. The other half (N = 16) experienced the countermand in the text and synthetic voice Data Comm display condition. A 30-second delay was implemented between the live countermand of the Data Comm message and the receipt of the message on the flight deck.

Before starting to fly, participants were familiarized with the touch-screen tablet and received a mini-flight briefing before each scenario (practice and experimental). Prior to flying the experimental scenarios, participants flew a 15-minute practice scenario in the Daytona Beach airspace. All scenarios were hand-flown. A notepad was provided for the pilot to use as necessary.

In all scenarios, a live controller (one of the voices heard on the party line) was available to respond to participants' questions while flying. No overlap occurred between the voice of the live controller and the synthetic speech. When the participant made an error, the live controller provided an appropriate, standardized reply to ensure that the participant was corrected back on course. Participants completed three surveys; one after each experimental scenario and a final usability survey after flying both scenarios.

RESULTS

First, do no harm…

Response times. Pilot response times were calculated and compared between Data Comm conditions to ensure that pilot actions were not delayed due to the presence of auxiliary synthetic speech relative to text-only Data Comm. Response times were analyzed separately for "Key Events," in which the pilots were required to make a change to the flight controls, and "Stability Events," in which pilots were only required to maintain flight precision. Three types of response times were analyzed for Key Events: 1) time to acknowledge the message from ATC; 2) time to initiate input to the flight controls; and 3) time to complete ATC instruction. For Stability Events, only time to acknowledge the message was analyzed.

Response times were calculated from message onset. Response times for both event types were compared by Data Comm type (text only vs. text+speech) using the Wilcoxon Matched-Pairs Signed-Ranks Test. With all Key Events combined, there were no significant differences in time to acknowledge the message, initiate inputs, or complete instruction (all $p > .50$). The ranges of achievable effect sizes (95% confidence intervals) confirmed the probability that there was no effect (a difference of zero; see Table 1). For Stability Events, pilots were significantly faster to acknowledge the message from ATC with text only than they were with text+speech, $Z = -19.46$, $p < .01$, presumably because with text+speech, they listened until the end of the message before responding. For Stability Events, the achievable effect size lay between .39 seconds and 2.65 seconds in favor of text only (with a range outside of zero, it is confirmed that this was likely a true, albeit small, effect of Data Comm condition, see Table 1).

Table 1. Median differences (Hodges-Lehmann estimator; see Bendre, 2010) and 95% confidence intervals (C.I.) for response times (sec) between text-only and text+speech conditions.

		Lower C.I.	*Median Difference*	Upper C.I.
Key Events	Acknowledge ATC Message	-2.56	*-0.49*^{ns}	2.02
	Initiate Input to Controls	-1.56	*0.08*^{ns}	1.65
	Comply with ATC Instruction	-5.13	*0.68*^{ns}	6.91
Stability Events	Acknowledge ATC Message	-2.65	*-1.66***	-0.39

^{ns} *non-significant, ** $p < .01$*
Negative values represent an advantage of text only and positive values represent an advantage of text+speech.

Similar call signs. The problem of responding to similar call signs should be alleviated with Data Comm, but auxiliary synthetic speech might minimize this benefit. The number of responses to similar call signs were counted and compared across Data Comm conditions using the Wilcoxon Matched-Pairs Signed-Ranks Test. The majority of pilots (24 out of 34) made no call-sign errors, and a comparison of the errors by conditions showed no difference, $Z = 0.00$, $p = 1.0$.

Countermanded clearances. When ATC countermands a Data Comm clearance via voice before it is displayed, pilots might be more likely to ignore the countermand when the Data Comm message is displayed both visually and via voice. A comparison of the total number of errors by condition shows six errors in the text+speech condition vs. four in the text-only condition, but this difference was not significant, $Z = -.63$, $p = .53$. The majority of pilots (22 out of 32 pilots) made no errors in either condition.

Pilot queries and live ATC interventions. Pilot queries to ATC and ATC interventions (e.g., to correct an error) may be indicative of communication difficulties. The number of pilot queries was compared by Data Comm condition using a Wilcoxon Matched-Pairs Signed-Ranks Test. Some pilots did query ATC (28 queries across both conditions); however the addition of synthetic speech did not affect the number of pilot queries to ATC, $Z = -.30$, $p = .76$. The majority of pilots (24/32) queried ATC equally in both the text-only and the text+speech condition. The number of ATC interventions was compared by Data Comm condition using a Wilcoxon Matched-Pairs Signed-Ranks Test. The number of interventions was not affected by the addition of synthetic speech, $Z = -1.89$, $p = .85$; and the majority of pilots (21/32) required interventions equally in both the text-only and the text+speech condition.

Second, help if you can…

Dwell time. The main hypothesis was that, compared to text alone, auxiliary synthetic speech would reduce the time pilots spent looking at the Data Comm display. Dwell time was measured two ways: 1) qualitatively, via post-scenario surveys, and 2) quantitatively, via video recordings of eye movement in the cockpit. Two coders, blind to the experimental condition, measured video-recorded dwell-time: one primary coder (whose data were used in the analysis) and a secondary coder (who recorded a subset of the data). The two coders were found to be in high agreement [$r(373) = .94$)]. In the post-scenario survey, pilots were asked to estimate the percentage of time they spent looking at the touch-screen display, at the instruments, out the window, or at "other." Paired *t*-tests compared the dwell-time estimates for each location by Data Comm condition. Results indicated that pilots in the text-only condition reported spending a significantly larger amount of time looking at the touch-screen tablet relative to pilots in the text+speech condition, $t(31) = 2.54$, $p < .05$ (no other comparisons were significant; all $p > .15$). The video measurements of dwell time revealed a similar pattern for total dwell time, with pilots spending significantly more time (an additional 20 seconds) looking down (in the direction of the tablet) in the text-only condition than in the text+speech condition, $t(31) = 2.50$, $p < .05$. It was possible, with 95% confidence, to find an effect as small as 3.75 seconds and as large as 36.92 seconds, in favor of text+speech (see Figure 2). A visual analysis of the data revealed that the 20-second advantage of text+speech seemed to be randomly distributed across communication events. In fact, it appears that much of the dwell-time advantage occurred between events, perhaps because pilots felt less of a need to check for messages with a speech cue in addition to the ding dong.

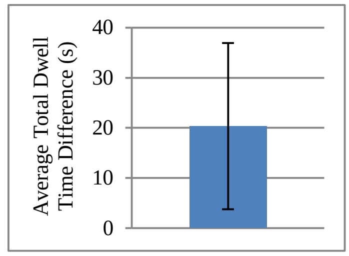

Figure 2. Average total dwell time advantage for text+speech Data Comm over text-only Data Comm. *Error bars are the 95% confidence intervals.*

Conditional clearance. When ATC issues a conditional clearance ("At Ormond, Climb to 3,000"), pilots may either forget to wait for the condition and respond early or forget to respond once the condition is fulfilled. The presence of a synthetic voice may mitigate this proactive memory lapse. A Wilcoxon Matched-Pairs Signed-Ranks Test was used to compare the number of errors by Data Comm condition. There were seven total errors in the text-only condition and two total errors in the text+speech condition. Only one pilot made fewer errors in the text-only condition and six pilots made fewer errors in the text+speech condition, a trend that approached significance, $Z = -1.89$, $p = .06$.

Pilot opinion

Results from the post-scenario questionnaires indicated pilots believed that communicative workload was low in both Data Comm conditions (with no significant difference, contrary to our hypothesis). Pilots also felt confident using the Data Comm system and believed that most people would learn to use it quickly. Pilots felt that both the text display alone and the computer-generated speech in addition to the text display were helpful, and equally so. Pilots were also in overall agreement that the computer-generated speech in addition to the text display was not distracting. Compared to communicating with live ATC, pilots did not prefer to communicate via text only, and they were undecided on a preference for text+speech over a live controller. When asked directly whether they agreed with the statement "I preferred communication with ATC using the text+speech display over the text-only display," pilots did show a preference for text+speech; a one-sample t-test found that the average agreement rating was significantly higher than the neutral scale anchor of "undecided," $t(31) = 2.06$, $p < .05$. This is in concert with the performance benefits found for text+speech.

DISCUSSION

The results suggest that an auxiliary synthetic speech display would offer several benefits without hindering pilot performance. Pilots in the text-only condition spent a total of

20 seconds longer looking at the touch-screen tablet relative to pilots in the text+speech condition during the approximately 30-minute flight—a result that is both statistically and operationally significant. Self-reported measures further corroborated this result. When pilots spend more time looking at the touch-screen tablet, this translates into less time spent looking out the window or at the instrument panel. The NextGen environment will increasingly compete for pilot's visual attention. For example, new traffic may appear on the Cockpit Display of Traffic Information while the pilot is communicating with ATC via Data Comm. Thus any additional time spent looking at the visual Data Comm display rather than at the instrument panel or elsewhere in the cockpit is operationally relevant.

The addition of the speech display may also have prevented pilots from acting early on a conditional clearance (perhaps because it provided an additional proactive memory cue). Fewer pilots acted early on these clearances or forgot to act on them when the condition was fulfilled with text+speech. No difference was observed in the number of pilot queries to ATC (e.g., requests for clarification, repeat instructions) or in the number of calls made by ATC to pilots (e.g., to correct an error).

Importantly, the auxiliary synthetic speech display did not appear to harm pilot performance. Pilots were not more likely to respond to similar call signs on the party line when communicating via the text+speech display compared to text alone. The implementation of synthetic speech did not induce pilots to ignore an earlier live voice countermand. Moreover, the addition of a synthetic speech display did not delay pilots' operation of the flight controls or compliance with ATC instructions. In some cases, pilots were faster to acknowledge ATC via the touch-screen tablet in the text-only condition, suggesting that pilots in the text+speech condition may wait to acknowledge the message until they have heard the full annunciated instruction, as they may do with live ATC. In certain situations, this may prove an advantage, preventing pilots from responding to an incomplete yet meaningful instruction early. However, in general, the addition of the synthetic speech display did not delay pilot response time. Lastly, participants tended to view the synthetic speech display favorably—it was deemed helpful and not distracting.

This proof-of-concept study has several inherent biases: The predominantly young participants presumably were proficient in texting and the design of the touch-screen tablet was ideal (actual Control Display Units on the FMS are smaller, shared with other applications, and often do not make use of color to code responses). Such biases may underestimate dwell time. Pilots also did not initiate Data Communications with ATC. The kneeboard plus touch-screen tablet solution presented in this experiment does not afford the ability to compose messages, and the implementation of such a capability may reveal itself as impractical. Here, no overlap occurred between the synthetic speech and the voice of the live controller. Controllers may contact the pilot while the synthetic speech is annunciating a Data Comm message and it is unclear how the two modes of communication would interact. Finally, as the concept was tested with a single pilot and tactical TRACON clearances, text+speech may have been given an advantage over text alone. Nonetheless, taken together, the results suggest that the text+speech display aided single pilot performance compared to a text-only display without introducing complications. A planned future study will evaluate the use of this technology in an en-route multi-crew flight deck environment that will remove some of the biases mentioned above and determine whether synthetic speech could interfere with simultaneous intra-crew and live ATC voice communications.

ACKNOWLEDGMENTS

This work was conducted in support of the Federal Aviation Administration (FAA), NextGen Advanced Concepts and Technology Development, Human Factors Division (ANG-C1). We thank our FAA program manager Dan Herschler for his insights throughout this work and Maura Lohrenz of the Volpe Center for her pertinent comments on an earlier draft.

REFERENCES

AT&T Natural Voices™ Text-To-Speech Engines: System Developer's Guide – Server, Server Lite, and Desktop Editions [Computer software manual] (2002). Retrieved July 5 2012 from http://www.naturalvoices.att.com/support/ATTNaturalVoicesTTS11.pdf

Bendre, S. M. (2010). *Nonparametric Statistics*. Unpublished manuscript, University of Hyderabad.

Bürki-Cohen, J. (1996). *An analysis of tower (ground) controller - pilot voice communications* (Report No. DOT/FAA/AR-96/19). Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Cardosi, K. (1993). *An analysis of en route controller-pilot communications.* (Report No. DOT/FAA/RD-93/11). Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Grayson, R. L., & Billings, C. E. (1981). Information transfer between air traffic control and aircraft: communication problems in flight operations. In C. E. Billings & E. S. Cheaney (Eds.), *Information transfer problems in the aviation system* (NASA Tech. Paper 1875). Moffett Field, CA: NASA Ames Research Center.

Helleberg, J. R., & Wickens, C. D. (2003). Effects of data-link modality and display Redundancy on pilot performance: An attentional perspective. *International Journal of Aviation Psychology*, *13*, 189-210.

Kerns, K. (1999). Human factors in air traffic control/flight deck integration: Implications of datalink simulation research. In J. A. Wise, V. D. Hopkin & D. J. Garland (Eds.), *Handbook of Aviation Human Factors* (pp. 519-546). Mahwah, NJ: Lawrence Erlbaum.

Lancaster, J.A., & Casali, J.G. (2008). Investigating pilot performance using mixed modality simulated data link. *Human Factors*, *50*, 183-193.

McCarley, J. S., Talleur, D., & Steelman-Allen, K. S. (2010). *Effects of data link display format and position on flight performance*. Univ. of Illinois Urbana-Champaign.

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

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multi-task pilot performance using advanced display technology. *Human Factors*, *45*, 360-380.