Ninth USA/Europe Air Traffic Management Research and Development Seminar (ATM2011)

CONTROLLER AND PILOT EVALUATION OF A DATALINK-ENABLED TRAJECTORY-BASED OPERATIONS CONCEPT

Eric Mueller, Dave McNally, Tamika Rentas, Arwa Aweiss, David Thipphavong, Chester Gong
NASA Ames Research Center
Moffett Field, CA, USA
Eric.Mueller@nasa.gov

Jinn-Hwei Cheng, Joe Walton, John Walker, Chu-Han Lee, Scott Sahlman
University-Affiliated Research Center
Moffett Field, CA, USA

Diane Carpenter
Science Applications International Corporation
Moffett Field, CA, USA

Abstract—**Results are presented for a pilot- and controller-in-the-loop evaluation of a 2016 timeframe datalink-enabled Trajectory-Based Operations concept with mixed voice and datalink operations. Eleven recently retired Fort Worth Center controllers and twelve current commercial pilots evaluated the concept over 28 hours of simulation time, providing quantitative metrics on controller workload and trajectory efficiency benefits and qualitative feedback on the feasibility of the concept. Eight experimental conditions were tested: four fleet-wide datalink equipage levels ranging from 0% (voice only) to 80%, along with two levels of traffic density. Feedback on the feasibility of the concept was positive from both pilots and controllers, though off-nominal conditions were not tested. The key objective finding of the evaluation was that controllers issued significantly more time-saving flight plan amendments under the datalink conditions than under the voice-only conditions, amounting to between five and twelve minutes of flying time savings per hour in the six-sector area tested. Statistically significant decreases in controller workload were also measured with increasing datalink equipage level, but this decrease was far smaller than the variation in workload across sectors and traffic density levels. The frequency with which aircraft requests for lateral trajectory changes were approved did not change with datalink equipage level for voice requests, but did increase for datalink requests; however, under all conditions, voice requests were significantly more likely to be approved than datalink requests.**

Keywords-trajectory based operations, datalink, air traffic control and management, human-in-the-loop simulation

I. INTRODUCTION

Air traffic delays in the United States' National Airspace System (NAS) will continue to increase as the number of daily flights increases unless overall capacity is expanded. Estimates are that the demand will increase up to twofold over the next fifteen years, with delays increasing much faster than that [1]. Currently, uncertainty in the future positions of aircraft leads to increased air traffic controller workload in maintaining safe separation between those aircraft, and airspace restrictions are put in place to control the flow of traffic to maintain workload at a reasonable level. These restrictions result in reduced airspace capacity and longer flight paths with higher fuel burn. A key concept of the NextGen air transportation system is the use of automation systems to track and predict the future positions of aircraft, termed Trajectory-Based Operations (TBO), which should lead to the removal of restrictions and increases in capacity and efficiency. The concept should also require minimal changes to flight deck equipage or controller systems.

Previous modeling and simulation studies of concepts similar to datalink TBO suggest that efficiency, capacity and safety benefits are realizable. Up to half a billion dollars could be saved each year through the use of integrated datalink and ground automation functions, primarily because capacity could be increased in controller workload-limited sectors and thereby reduce delays [2-4]. Trials of datalink in the Miami Air Route Traffic Control Center (ARTCC) between 2002 and 2003 indicated significant reductions in voice frequency congestion were possible with datalink alone (no TBO required), a problem that is driving European airspace users to equip with datalink more quickly than US users [5]. Safety could also be improved by reducing the frequency of operational errors: out of a representative set of 58 errors analyzed in a recent study, 20% were the result of pilot-controller misunderstandings leading to missed altitude level-offs, and 52% were from improperly analyzed and issued horizontal or vertical trajectory changes [6]. Instances of errors in either category should be reduced with an integrated datalink-TBO concept.

While the benefits of datalink usage are generally widely recognized and datalink is thought to be a key enabling capability for TBO, translating this perception into a convincing cost/benefit analysis has been an elusive goal for two major reasons. First, the principal direct benefit of datalink appears to be the reduction of controller workload resulting from fewer voice transmissions, most of which are for transfer of communications [2, 7, 8]. However, this benefit mechanism is difficult to monetize without broaching the sensitive subjects of either reducing the number of air traffic controllers or increasing the allowable number of aircraft in a sector [9]. The second problem with producing a solid cost/benefit analysis for datalink is that costs to airspace users (in terms of additional equipage) are quite high under most concepts but the direct benefit to equipped aircraft is no more than that for unequipped aircraft [10]. This creates an incentive for each user to wait to equip as long as possible unless artificial requirements (i.e. regulations) are imposed to provide better service to equipped aircraft. What is needed is a way to equate lower controller workload with direct efficiency, capacity and/or safety benefits for the system generally and, more preferably, for equipped aircraft.

This paper reports the results of a high fidelity human-in-the-loop (HitL) simulation evaluation of a particular near-term concept, envisioned as a first step towards en route TBO. "Near-term" is defined as the time horizon in which "equipped" aircraft need nothing new beyond currently-flying (2010) advanced flight deck systems, existing datalink (VDL-2) is used, and existing controller tools are improved, moved to the R-side, and integrated with datalink. A detailed description of that concept and a non-HitL simulation analysis of its potential benefits may be found in [11]. The primary goal of the evaluation was to demonstrate the feasibility of TBO in a mixed voice and datalink communication modality using several new tools on the controller's R-side display but no new aircraft equipage. The second goal was to explore and objectively measure benefits provided by these new tools when integrated with datalink in terms of controller workload and aircraft flight efficiency. The ultimate goal is to provide results to facilitate the design of a cost-effective concept for allowing all aircraft to take maximum advantage of TBO.

II. METHOD

A. Participants

Three teams of commercial pilots and recently retired air traffic controllers participated in the simulation over four weeks in September and October 2010. Each team consisted of four test controllers to work traffic, one staff controller to help coordinate and train the test controllers, two flight crews (four pilots) to fly the two high-fidelity cockpit simulators, one staff pilot to coordinate, train and observe the flight crews, and eight pseudo-pilots to fly the lower-fidelity desktop cockpit simulators.

To ensure feedback on the feasibility of the concept would closely match results from an operational evaluation, a pool of operators and users of the air traffic system was recruited. The eleven air traffic controllers had worked Fort Worth Center airspace operationally for an average of 25.5 years, and had been retired on average 2.3 years (maximum of 6 years retired). The six flight crews—twelve pilots—who flew the cockpit simulators averaged 12,900 hours of flight time on a variety of commercial transport aircraft and were all currently employed by a major airline. The three crews who flew the 747-400 simulator cab were rated on that aircraft and had considerable CPDLC[1] experience; the other three crews flying the research simulator cab were rated on 757 or 767 aircraft and only had ACARS[2] datalink experience. This difference was intended to test the difficulty with which pilots new to CPDLC but familiar with the Flight Management System (FMS) interface would be able to adapt to the new communication modality. The eight pseudo-pilots who flew the desktop cockpit simulators were either currently-rated commercial pilots or retired controllers, and five of the eight had participated in NASA experiments as pseudo-pilots before.

B. Experiment Design

The experiment design was guided by the key assumptions in the target near-term TBO concept [11], and by two important variables relevant to TBO: percentage of aircraft equipped with integrated datalink/FMS and overall spatial density of traffic. Three key equipage assumptions for the concept were made: ground-based trajectory automation was integrated with the Center radar (R-Side) controller position; mixed equipage datalink and voice operations would occur in the same airspace; and Future Air Navigation System (FANS-1/A) integrated FMS/datalink would be used on "equipped" aircraft. The values of the experimental variables are shown in Table I. The first team of pilots and controllers evaluated every cell of that matrix three times, while the second and third teams had time to evaluate each only twice.

Four levels of fleet-wide datalink equipage were tested: 0%, the baseline against which the other conditions are compared; 20%, the expected domestic US FANS-1/A equipage level when the datalink infrastructure is enabled in 2016; a 50% equipage level consistent with 2025 projections for NextGen; and an 80% level, a long term estimate of the percentage of aircraft for which a reasonable business case might be made for upgrading to FANS-1/A.

Two levels of spatial traffic density were tested. The first level represented current traffic density in the southern US. The second traffic level was about 50% higher than this nominal density, a value selected because a major benefit of increased R-side automation and datalink is expected to be reduced controller workload. Density in most scenarios tended to increase with time so even the high density recordings had periods of relatively light traffic; because of this within-scenario variation the high and low traffic density conditions are not compared directly and instead the directly measured workload of the controller is correlated with other metrics.

TABLE I. EXPERIMENT MATRIX

Traffic Density	**Datalink Equipage**			
	0%	***20%***	***50%***	***80%***
Nominal (1x)				
Very Busy (1.5x)				

C. Traffic Scenarios

The test airspace included six high-altitude sectors in eastern Fort Worth Center (ZFW); this area was chosen because of its good mix of arrival and departure flows from Dallas-Fort Worth and Houston International Airports, and the over-flight traffic that intersects with these arrival and departure flows. The six sectors were combined into three, outlined in blue in Fig. 1, with a single controller working each of the combined sectors. Operationally, when traffic levels are low, sectors 90 and 71 are frequently combined, while it is rare for the other sectors to be combined. However, the enlarged area of responsibility provided an additional level of workload

---

[1] Controller-Pilot Data Link Communications

[2] Aircraft Communications Addressing and Reporting System

with which to stress the controllers and determine their limits under the concept. A single "ghost" controller managed traffic outside the three high-altitude test sectors.

Traffic recordings were generated using en route Center radar track and flight plan data recordings from the FAA's Fort Worth Air Route Traffic Control Center (ARTCC) from July 29 and 30, 2010. For the "current day" (1x) traffic density recordings, the flight plans and initial positions of aircraft were used to initialize the aircraft simulators described below; once a scenario started, those simulators calculated and updated the aircraft positions. For the very busy scenarios (1.5x), traffic recordings from the same two days were combined with individual flights from other times and dates until the average traffic count was 50% higher than the "current day" recordings. These individual flights were time-shifted as needed to limit the number of short term conflicts (<4 minutes to loss of separation) at the start of the scenario. Special attention was paid to selecting flight plans for the high fidelity aircraft cab simulators (described below) that would include traffic conflicts to ensure interactions between the controllers and the flight crews.

The datalink-equipage scenarios were generated by taking the six traffic recordings (three 1x, three 1.5x) and assigning appropriate types of aircraft, either datalink (e.g. 777s but not 727s) or voice equipage depending on the required percentage. To allow a direct comparison between the recordings at different datalink levels, every recording was used to create four scenarios – one for each datalink level. The aircraft IDs were randomized to hide the fact that recordings were being reused and controllers were rotated from sector to sector so that they did not see the same recording when they were working the same sector.

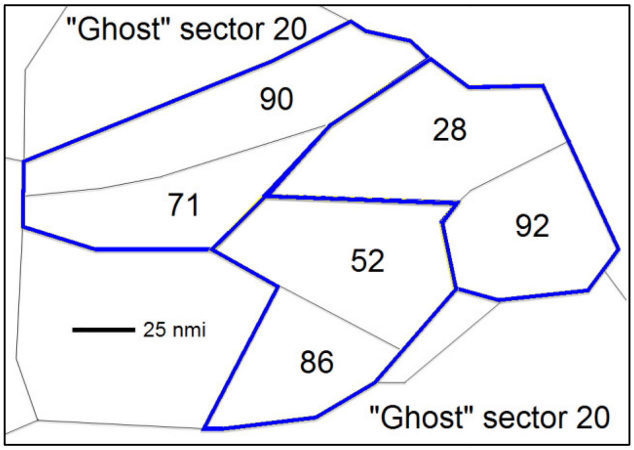


Figure 1. Three test sectors in Fort Worth Center

D. Simulation Facilities

The simulation facilities included four emulated Center radar (R-Side) controller positions, two high-fidelity cockpit simulators, and four pseudo-pilot stations each capable of simulating 15 or more aircraft.

The emulated R-side controller positions were driven by the Center-TRACON Automation System (CTAS) trajectory analysis software developed at NASA Ames Research Center and tested in numerous research and operational settings [12-15]. The controller interface was based on a standard keyboard and mouse, not the R-side keyboard and trackball normally used, and datalink-equipped aircraft were colored green to distinguish them from the white voice-only aircraft. Controllers were primarily tasked with avoiding conflicts and managing handoffs; they did not have to meter arrival traffic, implement traffic flow management initiatives or deal with weather. The four controller positions used in this experiment are shown in Fig. 2. The new capabilities available to the controllers with this system are described in the next section.

The experiment employed high-fidelity cockpit simulators to enable realistic pilot-controller interactions, including realistic datalink response times, flight deck procedures, and the conditions under which the pilots and controllers would abandon datalink and use voice to communicate or negotiate a trajectory change. A picture of the inside of one of the cabs, a 747-400 Level D simulator, is shown in Fig. 3. This simulator contains actual FANS1/A hardware and is as true to real aircraft as is possible on the ground [16]. The second cockpit is similar to that shown in Fig. 3, but uses a 737-style FMS coupled to a 757-like aerodynamics model and is designed to allow modification for research purposes. Pilots were given specific procedures to follow in responding to datalink clearances, procedures that were developed and refined in earlier experiments in the same facility [17]. These procedures were used in the 20%, 50% and 80% datalink conditions, and voice procedures were used in the 0% datalink conditions.

Figure 2. Emulated controller R-side stations

Figure 3. 747-400 cab simulator

All other aircraft in the scenarios were controlled by one of four pseudo-pilot stations [18]. Each station managed all the aircraft in a particular sector and was staffed by two professional pilots. One pilot handled voice communications with the sector controller, while the other pilot implemented the controller's instructions using a simulated cockpit display. Datalink instructions were automatically accepted and executed according to a distribution of response time delays measured in a previous pilot-in-the-loop experiment [17]. The median response delay for a lateral flight plan amendment was 56 seconds (minimum 10, maximum 165) and the median delay for a vertical amendment was 39 seconds (minimum 10, maximum 85). Each pseudo-pilot station generally handled between 5 and 20 aircraft at a time, the same number the controller was managing.

E. Controller Automation Tools

The primary functional improvements of the emulated R-side system over today's R-side displays were the inclusion of an interactive, rapid-feedback conflict detection and trial planning capability (a form of which is available only on the controller's D-side station) [19, 20], a conflict resolution advisory function based on the Advanced Airspace Concept [21-23], and an integrated datalink capability to automatically construct and transmit the flight plan amendments generated by the controller using the trial planning capability [24]. The Flight Data Blocks (FDB) on the controllers' displays, shown in Fig. 4, contained the usual fields found on an R-side display plus enhancements. Additional fields indicated the destination airport (KDFW), the current sector ownership (86), time-to-loss of separation (LOS) in minutes (6), a conflict resolution advisory message (L20), and an aircraft datalink request indicator (C360) when applicable. Controllers initiated the route trial planning function by clicking on the destination airport field, and triggered the altitude trial planning function from the altitude field. Clicking the time-to-LOS field in the FDB brought up a graphical depiction of the aircraft conflict geometry, as shown by the red lines in Fig. 5, and clicking the autoresolution advisory field displayed the recommended conflict resolution maneuver as a trial plan, shown by the yellow line. The datalink message is automatically updated as the controller manipulates the trial planner, either by moving the auxiliary waypoint or selecting a different return waypoint. The message is transmitted to the aircraft and amended in ATC "Host" computer database when the controller hits the "Amend/Uplink" button in the trial plan window, shown in Fig. 6. The trial plan window also contained a list of fixes at which the trial plan should rejoin the original flight plan, the actual datalink message to be transmitted, and buttons to simply amend a clearance without sending a datalink message, reject a pilot request, or revert to a previous flight plan. The datalink messages, conforming to the CPDLC standard [25], were formatted to be acceptable to the simulators in this experiment or to real FANS-equipped aircraft.

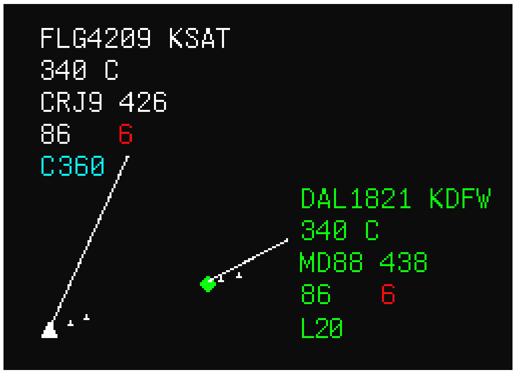


Figure 4. Flight data blocks with conflict notification, resolution advisory and datalink request from aircraft

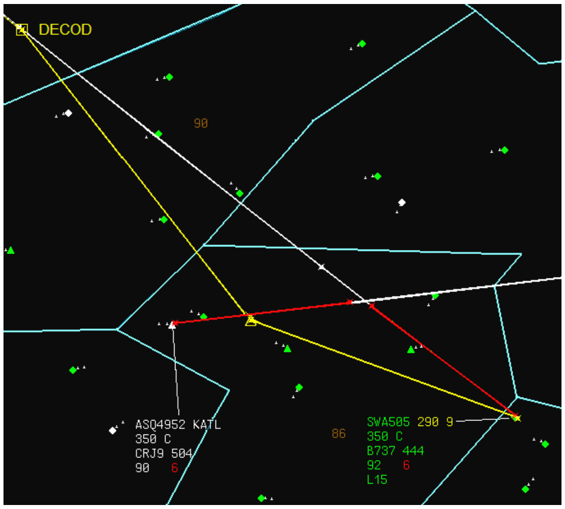
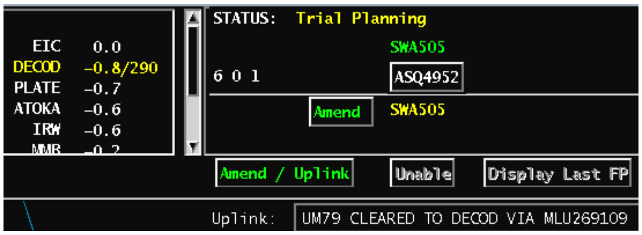


Figure 5. Conflict geometry (red) with trial plan resolution trajectory (yellow)


STATUS: Trial Planning

EIC 0.0
DECOD -0.8/290
PLATE -0.7
ATOKA -0.6
IRW -0.6
MMB -0.2

Uplink: UM79 CLEARED TO DECOD VIA MLU269109

Figure 6. Trial plan and CPDLC interface window

F. Training

Each team participated in the experiment for four days, and approximately 1.5 days were spent training with specially designed scenarios in CTAS to minimize the degree to which novel aspects of the interface compromised evaluation of the functional aspects of the R-side automation. Before progressing to the data collection scenarios a checklist of functions was reviewed with each controller to ensure consistent use of the automation. Training for all the pilots was accomplished considerably faster than it was for the controllers; there were far fewer changes to current procedures for the pilots, and all pilots were able to master the datalink procedures in a matter of hours regardless of CPDLC experience. Training for the pseudo-pilots on procedures and phraseology was conducted for a week before data collection to ensure they were capable of managing up to twenty aircraft at a time. Pseudo-pilots were able to attain a level of proficiency in that first week that was consistently high during the data collection weeks.

G. *Simulation Procedures*

Scenarios were presented to the participants in a pseudo-random order intended to evenly distribute the high and low workload scenarios and the datalink-equipage levels, and to minimize any lingering learning-curve effects. In each block of four scenarios two were 1x density and two 1.5x. All four datalink levels were represented. Within those constraints the order was randomized. The order of the scenarios between teams was selected to ensure particular scenarios weren't overrepresented at the beginning or end of data collection.

To investigate the effect of datalink-enabled TBO on the approval of flight deck requests, the pilots in the cab simulators were instructed to ask for route efficiency opportunities whenever they arose. A simulation engineer passed direct routing opportunities to the appropriate pseudo-pilot position every 5 minutes. The opportunities were for either voice or datalink aircraft, and the appropriate modality was then used to relay the request to the controller.

III. RESULTS

The following subsections will describe subjective feedback on the feasibility of the TBO concept and objective benefits measured during the simulation.

A. *Feasibility of TBO Concept*

The primary goal of this evaluation was to determine whether it is feasible to use both datalink and voice communications in the same airspace to convey routine messages, simple route, and altitude amendments and complex routing changes. It was the overwhelming consensus of the pilots and controllers that such mixed-equipage operations would be appropriate for most of the domestic en route airspace. Exceptions to this include the very busy northeast regions of the US, particularly Cleveland, Indianapolis, Washington and New York centers.

Controllers provided generally positive comments about the R-side automation *functionality* additions, though they did have suggestions on ways to improve the automation *interfaces*. Controllers were observed to move from predominantly using the trend vectors to predict future positions of aircraft to using the trial planner; controllers reported that the trial planner provided significantly more information than the trend vectors because of its use of flight plan intent and aircraft performance models. Improvements to the vertical trajectory modeling were suggested, including using the current altitude rate to correct the predicted rate from the performance model. The autoresolution advisories frequently corresponded with what the controller would have decided to do, but straightforward improvements like altitude-for-direction-of-flight conventions and taking into account the arrival status of aircraft would have improved its performance. Controllers reported they relied exclusively on the advisories only when extremely busy, and otherwise it did not lower their workload because they had formulated a resolution maneuver already. The ability to transmit a datalink message automatically constructed from the trial planner was frequently cited as a major advantage over current operations, though observation of the controllers' actions and their comments suggest that much of the apparent reduction in workload resulting from datalink usage was from not having to handle transfer of communications over voice. This finding is consistent with past studies of datalink TBO [2].

Controllers and pilots both expressed concerns stemming from datalink response delays and particularly the lack of an immediate response that at least acknowledges receipt of a datalink message. Uplink messages sent to existing flight decks take a median of one minute for pilots to load, visualize or otherwise evaluate, execute, and send a response, whereas controllers reported that the maximum allowable time (rather than the median time) needs to be less than sixty seconds. This closely matches the opinions of controllers using datalink in European trials [26]. This response delay meant that controllers, on average, were comfortable using datalink to issue resolutions when the conflicts were at least 4.5 minutes away for altitude changes and 6.4 minutes away for lateral path changes. Otherwise, controllers would resort to clearances issued by voice. Downlink messages requesting a route or altitude change were displayed in the flight data block, but were sometimes not noticed by the controller. This could be a training issue that would not appear in practice; however the post-run discussions on this topic revealed the desire for some kind of acknowledgement capability through messaging or voice procedures, or simply faster response times via better interfaces or datalink procedural changes.

The pilots found little else to object to in the concept, likely a result of requiring no new flight deck equipage and only minor modifications of existing datalink procedures. All pilots quickly learned to use CPDLC regardless of prior experience, calling it "simple" and "straightforward." Certainly the use of a new communication modality like datalink will introduce new considerations regarding human-machine interaction errors [17, 27, 28], but solutions to these problems have been identified in prior research and the problems are unlikely to be fundamental limitations that preclude wide use of datalink.

Perhaps the most important qualitative result of the simulation was that a mix of voice and datalink communications appears to be the best way for controllers and pilots to interact. Voice communications are critical in situations when time-to-LOS is less than several minutes, clarification is needed because several datalink messages have been exchanged, a complex reroute needs to be negotiated, a pilot needs to deviate due to weather or aircraft performance, or in emergency circumstances. Datalink communications are most effective when the instruction is routine and the expected response is "WILCO," or when a more complex instruction can be automatically loaded and executed by the aircraft's FMS. Based on this study's findings, there is no reason to suppose a future ATC system must or should attempt to completely replace voice with datalink communications.

B. *Benefits of TBO Concept*

The following sections will present specific objective benefits of the TBO concept. Note that, in comparing the distributions of ratings using an analysis of variance

(ANOVA), p-values—the probabilities that two data sets have equal means—displayed at the bottoms of plots are always relative to the 0% group; unless otherwise noted the 20%, 50% and 80% groups are not compared with each other. The measure of "statistical significance" used here is p=0.10, which is less strict than the usual p=0.05 because data collected in a HitL is necessarily sparse, the presence of other experimental factors creates large variations in the variable being tested, and because subjective feedback (e.g. in terms of workload versus datalink percentage) agrees with the statistical tests.

1) Flight Plan Amendments

One of the most important proposed benefit mechanisms of a datalink-enabled TBO concept is that additional automation would reduce controllers' workload for a given level of traffic, which in turn would allow controllers to provide more efficient conflict resolution maneuvers and respond positively to aircraft requests more frequently. This hypothesis can be tested broadly by examining the predicted change in flying time for every flight plan amendment input by the controller. A histogram showing the distribution of flying time changes for every flight plan amendment in the experiment is given in Fig. 7. That figure provides context for the types of trajectory changes used by controllers en route that can impact efficiency, and indicates the potential for time savings or delays from each type of amendment. Fig. 7 shows that direct-tos (D2s), lateral amendments in which an aircraft flies direct to a downstream waypoint bypassing at least one waypoint on its route of flight, account for the majority of flying time savings for lateral route changes. Flying time differences for those maneuvers take into account the effect of the wind field, and the controller display includes the predicted flying time difference for the D2 route or any route change. Route amendments, in which at least one new auxiliary waypoint is added to the flight plan of the aircraft, usually add flying time. In both of these cases a flying time savings should translate into a fuel savings as well, although the actual fuel savings was not calculated during the evaluation. Vertical trajectory changes showed the largest range of flying time differences: from ten minutes of savings to six minutes of delay. This large variation arises because the airspeed is predicted to change with altitude; however the fuel burn will be based on the aircraft's optimum altitude for efficiency rather than simply the flying time.

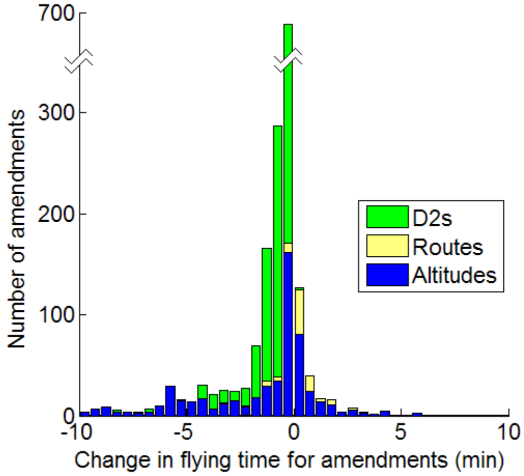


Figure 7. Distribution of flying time changes for flight plan amendments

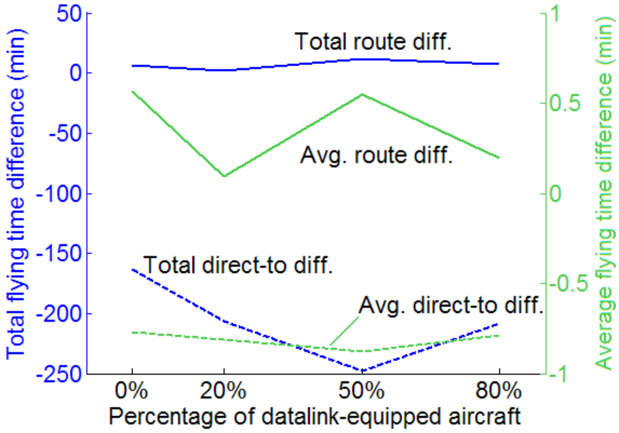


Figure 8. Total and average flying time differences for route and D2 amendments

The *key objective finding* of this experiment is that controllers issue more time-saving flight plan amendments when more datalink-equipped aircraft are under their control; the relationship between these variables is shown in Fig. 8. The dashed lines indicate results for D2 amendments while solid lines are for general route amendments. Blue lines show the total flying time differences for all amendments at a particular datalink percentage while green lines represent the average flying time difference for a single amendment.[3] The data indicate that at most a weak relationship exists between the average flying time savings for a single amendment and overall datalink equipage level (dashed green line), with the average D2 savings reaching a maximum of 0.88 minutes at 50% equipage compared with 0.78 minutes at 0% and intermediate savings at 20% and 80%. The key variable that *does* depend on datalink percentage is the number of flight plan amendments, specifically D2s, that are sent at higher equipage levels: because D2s save flying time the increased number of these amendments means a greater overall savings at higher datalink equipage levels (dashed blue line). The savings are significant, rising from 163.5 minutes at 0% to 207 minutes at 20% and 80%, and topping out at 248 minutes for 50% equipage, each over a total of seven hours of scenario time. These represent, respectively, a 27% and 52% increase in savings over the baseline voice-only condition. These results are based on analysis of 1104 flight plan amendments (1013 of which are D2s) over 28 hours of simulation and are summarized in Table II. If confirmed in operational trials, this result would indicate an additional five to twelve minutes of flying time savings per hour could be realized in just the eastern portion of ZFW as tested here. It should be reiterated that the same traffic recordings were used for the four datalink equipage levels (though the call signs were changed) so the comparison is direct, and that the additional savings occurs above what would be expected in voice-only conditions. An analysis of the percentage of aircraft that spent time off their flight plans shows that for all conditions it was between 7.43 and 7.95%, so it is not the case that controllers in the voice-only condition still issued time-saving D2s but did not enter them as flight plan amendments. The experimental finding that more time-saving

---

[3] Flying time difference = new flight plan trajectory flying time minus old flight plan trajectory flying time

amendments are sent at higher datalink equipage levels represents a significant and direct benefit mechanism for airspace users.

TABLE II. D2 FLYING TIME SAVINGS

D2 Variable	**Datalink Equipage**			
	0%	***20%***	***50%***	***80%***
# Amendments	212	255	282	264
Average Δt (min)	-0.77	-0.81	-0.88	-0.79
Total Δt (min)	**-163.5**	**-206.8**	**-248.0**	**-208.0**

2) *TLX Workload Ratings*

The result from the previous section that higher datalink equipage levels result in more time-saving amendments being issued raises the question of whether it is a general reduction in controller workload that accounts for the additional amendments, whether it is due to the ease with which amendments can be sent to the Host and to datalink-equipped aircraft with a single mouse click, or something else. End-of-run workload ratings were provided by controllers using the NASA Task Load Index (TLX) scale [29], and those ratings indicate that at least some of the difference is due to workload reductions at higher equipage levels.

The TLX workload scale asks participants to rate their overall workload during a particular scenario along six dimensions: mental demand, physical demand, temporal demand, performance, effort and frustration. These component ratings can then be examined separately or averaged together to gain an overall picture of a participant's workload during a single scenario. In order to capture and analyze the peak workload during each scenario, participants were asked to provide ratings corresponding to the busiest periods they experienced. The ratings were normalized according to z-score for each participant, by which the mean TLX rating for a given participant became a zero z-score, and for every standard deviation in TLX rating away from a given participant's mean rating the z-score was higher or lower by one point. This transformation allows a more direct comparison across the different participants: while they were given guidelines on what each rating meant, inevitably individual controllers will have different average ratings and use more or less of the full scale. Workload ratings for most scenarios and sectors were low and only about 17% of the scenarios were given TLX average ratings of 5 or higher.

Fig. 9 shows the average of the six component ratings for each participant in each scenario—labeled "single run"—broken down by datalink percentage. An analysis of variance was performed and p-values are included in Fig. 9. The figure shows a likelihood of between 9 and 12% that workload at 50% and 80% datalink equipage is equivalent to the workload at 0% (p=0.094 to 0.118). Controllers reported anecdotally that they believed workload was much lower when datalink equipage was higher, so the relatively poor p=0.118 confidence level of a measurable decrease in workload is likely due to the wide variation in workload ratings for a particular datalink equipage level—not because the difference is due to chance. This variation occurs because different sectors and different traffic recordings had different traffic densities, but each point in one of the four equipage levels (e.g. sector 86, 1.5x traffic density, 80% equipage) corresponds to the exact same point in one of the other equipage levels (e.g. sector 86, 1.5x density, 0% equipage). The important result from Fig. 9 is that for the same traffic densities and sectors, a datalink equipage level of 50% or more provides a measurable reduction in controller's self-rated workload.

The individual workload components of temporal demand, effort and, interestingly, physical demand experienced a statistically significant decrease (at the p=0.10 level) with increasing datalink equipage. No statistically significant results were found for the other three categories (frustration, mental demand, performance). The ratings for temporal demand only become significant when 80% equipage is reached. The effort ratings are significant at 50% equipage (p=0.070) but not at 80% (p=0.152). The physical ratings are significant at both 50% and 80% equipage (p=0.069 and 0.038 respectively), perhaps because of the very significant reduction in the amount of talking required of the controller (see the final section in Results).

The end-of-run TLX ratings provided by the controllers suggest, consistent with their verbal feedback, that datalink equipage levels of 50% or higher result in measurably lower workload than the baseline voice-only condition. Together with the results of the previous section it appears that higher datalink levels reduce controller workload and allow controllers to send more time-saving flight plan amendments to both voice and datalink aircraft.

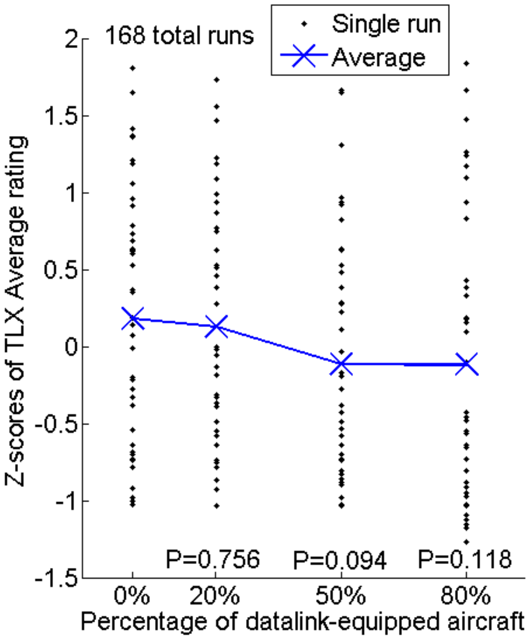


Figure 9. TLX Average workload ratings by fleet datalink percentage

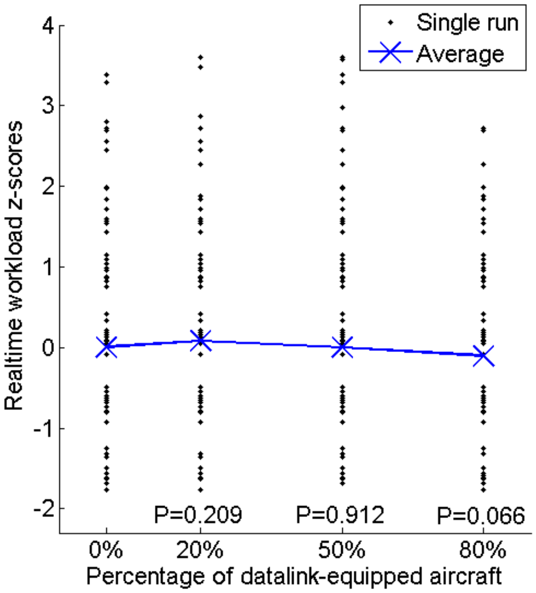


Figure 10. Realtime workload z-scores as a function of fleet datalink percentage

3) *Realtime Workload Ratings*

In addition to end-of-run TLX workload ratings controllers provided "realtime" ratings of generic workload (i.e. not categorized into mental, physical, etc.) every two minutes during the scenarios. Such a realtime measure is useful because controllers select a workload rating quickly without "over-thinking" the metric, it does not rely on controllers' memory of earlier workload, and it allows for comparison of metrics at specific times within scenarios rather than scenarios as a whole. The realtime workload ratings are normalized into participant-specific z-scores in the same way the TLX scores were normalized; for reference, the mean of the raw scores was 2.8 out of 7, and only 9.7% of the ratings were 5 or higher.

A variance analysis of the realtime workload data given in Fig. 10 shows a significant (p=0.066) reduction for the 80% case compared with the voice-only case, mirroring the results from TLX, but actually indicate that workload reaches a peak at 20% datalink equipage. This last result was not reflected in the TLX data and was not consistent with controllers' comments (controllers suggested that the higher the proportion of datalink aircraft the better, and 20% datalink was preferable to voice-only), so it may simply be due to chance. The realtime workload data appear to confirm the TLX results that workload decreases with the highest datalink equipage levels, but further study is required to determine whether a small percentage of datalink-equipped aircraft will actually increase a controller's workload.

A slightly different view of the realtime workload ratings is provided by Fig. 11. That figure shows the cumulative proportion of workload ratings below a given z-score for each datalink percentage. Lines that lie to the left of and above other lines represent conditions with lower workload because a higher percentage of the workload ratings in that condition are below the z-score on the x-axis. The take away from this figure is that there is a lack of very high realtime workload ratings at the 80% datalink level. This may indicate that at low traffic levels the 80% datalink condition does not have a major effect on workload, but that at the highest traffic densities workload does decrease if more aircraft are datalink equipped. If borne out by further study, such a finding would suggest datalink-enabled TBO is most useful when large numbers of aircraft are managed by a controller and that the maximum number of controllable aircraft may be higher than it is with today's systems.

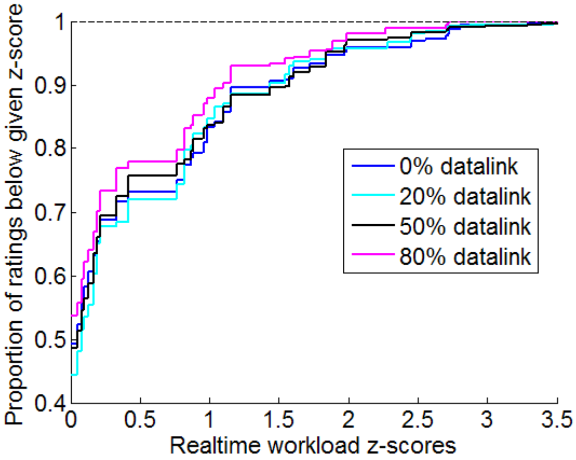


Figure 11. Cumulative distribution of realtime workload ratings by fleet datalink percentage

4) *Aircraft Requests*

A hypothesized benefit mechanism of datalink-enabled TBO is increased acceptance of aircraft (generically, "user") requests with higher datalink equipage levels because of reduced workload and the ease with which a datalink request could be evaluated and approved without verbal pilot-controller exchanges. As specified previously, D2 requests were made by the pseudo-pilots approximately every five minutes according to test procedures, and once or twice per scenario by the high-fidelity cockpit simulators; the total number of requests was relatively constant among the datalink equipage conditions (ranging from 55 to 70) and were generally distributed between voice and datalink in the same proportion as the overall fleet equipage percentage.

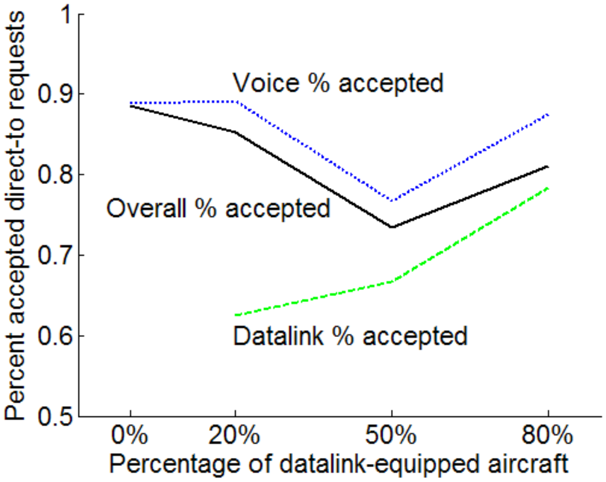


Figure 12. D2 request acceptance rate

There is no significant trend towards increased acceptance of simple route requests with increasing datalink equipage levels. Fig. 12 shows the trend in overall percentage of approved requests, approved voice requests and approved datalink requests. The overall and voice request approval rates are roughly 85 to 90% under most conditions, but drop to about 75% approval for the 50% equipage level. No explanation for this minimum is apparent. In contrast, the approval rate for datalink requests increases from 63% to 78% as fleet datalink equipage increases. This is likely partly a result of training issues, in which at lower datalink equipage levels controllers are not scanning flight data blocks for aircraft requests as often as they are at high equipage levels, and partly a result of workload. Because workload tends to decrease with increasing datalink equipage the controller has more time to look for D2 requests when more aircraft are datalink-equipped. The voice requests do not show this trend because controllers were forced to immediately respond to such requests, while they could intentionally or unintentionally ignore datalink requests. Perhaps the only unequivocal take-away from this plot, and one that was unanimously confirmed by controller feedback, was that voice requests are far more likely to be approved than datalink requests. This may suggest datalink should be used for routine operations that will reduce voice frequency clutter, like transfer of communications, in order to free the voice channel for aircraft requests and other negotiations that are more difficult to accomplish with textual datalink. It should be emphasized that these results apply only to *simple* route requests that can be made either by voice or datalink; more complex route requests not studied here would only be possible with datalink. This benefit mechanism has not been quantified.

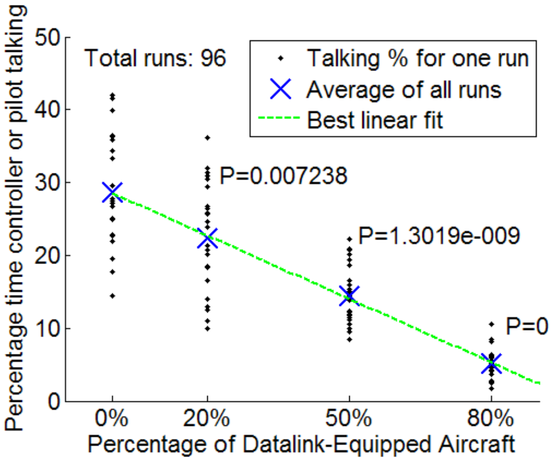


Figure 13. Relationship between fleet datalink percentage and voice usage

5) *Voice Channel Analysis*

The final set of analyses presented here relates to improved modeling of controllers' voice communications patterns under the datalink-enabled TBO concept rather than measurement of a particular benefit mechanism. Fig. 13 shows the relationship between the fleet datalink equipage level and the percentage of time either the controller or pilot was talking on the voice frequency in that particular sector. As in the previous results divided by fleet equipage, there is large variability within any single condition because different sectors and traffic recordings have widely differing numbers of aircraft. What is interesting about Fig. 13 is not the decrease in use of voice with increasing proportion of datalink equipage—that result was inevitable—but that the average talking times for the different conditions are very nearly linear with fleet datalink equipage. The talking times ranged from 28.5% when all aircraft are voice-equipped to almost exactly zero when all aircraft are datalink-equipped. The best linear approximation to this relationship is

$$\text{Mean Talking \%} = -0.29 * \text{DL\%} + 28.5 \qquad (1)$$

IV. CONCLUSIONS

The results of a pilot- and controller-in-the-loop evaluation of a datalink-enabled Trajectory-Based Operations concept are presented. The concept was designed for near-term implementation and therefore supported a mixture of voice- and datalink-equipped aircraft in the same airspace, currently flying aircraft Flight Management Systems integrated with datalink, and a set of new R-side controller automation tools. Four levels of fleet-wide datalink equipage were tested, ranging from voice-only operations up to 80% equipage, along with two levels of traffic density.

The most important objective finding of the evaluation was that controllers issue more time-saving lateral route amendments under higher datalink equipage levels than they do under voice-only operations, and the total resulting time savings is between 27% and 52% more than the baseline case. This translates into between 5 and 12 minutes of flying time savings per hour in the six-sector area of eastern Fort Worth Center tested. Realtime measures of controller workload and end-of-scenario evaluations indicated that increased datalink equipage levels reduce workload, but the reduction was small in comparison to the variation due to the number of aircraft in the sector. Pilots reported they were very comfortable using the datalink procedures provided to request flight plan amendments or approve controller clearances. Controllers recommended numerous improvements to the simulated R-side system, but indicated that the functionalities tested here could significantly improve the ease and efficiency with which they managed traffic once the functionalities were integrated into an operational system. It was found that *simple* aircraft route requests were significantly more likely to be approved when requested by voice than by datalink, and that the percentage of datalink-equipped aircraft did not have a clear effect on the frequency with which all requests were approved, though it did have a positive effect on the frequency with which datalink requests were approved.

REFERENCES

[1] Next Generation Air Transportation System Integrated Plan," Department of Transportation, Dec. 12, 2004.

[2] H. Hasan, K. Leiden, S. Mondoloni, D. Kozarsky and S. Green, "An Initial Benefits Assessment of Distributed Air/Ground Traffic Management Concept Elements," AIAA 2003-6806, AIAA Aviation Technology, Integration and Operations Conference, Denver, CO, Nov. 2003.

[3] M. Bennett, D. Knorr and J. Rakas, "Economic Benefits of Increased En Route Sector Capacity from Controller-Pilot Data Link Communications," 83rd TRM Annual Meeting, Transportation Research Record, 2004.

[4] Computer Sciences Corporation, "Life-Cycle Cost/Benefit Assessments of DAG-TM Concept Elements 5, 6 and 11: Version 4.0, Final Report,"

Air Traffic Management System Development and Integration, NASA Ames Research Center, DOC ID CSC/ATMSDI/CTO-08-10/040630, June 2004.

[5] R. Bolczak, J. Gonda, W. Saumsiegel and R. Tornese, “Controller-Pilot Data Link Communications Build 1 Value-Added Services,” Digital Avionics Systems Conference, Salt Lake City, Oct. 2004.

[6] R. Paielli and H. Erzberger, “Improved Conflict Detection for Reducing Operational Errors in Air Traffic Control,” AIAA 2004-6392, AIAA Aviation Technology, Integration and Operations Conference, Chicago, IL, Sept. 2003.

[7] R. Bolczak, “Operational Evolution of an Integrated URET/CPDLC Capability,” 21st Digital Avionics Systems Conference, Irvine, CA, Oct. 2002.

[8] J. Celio, “CPDLC Integration with URET – A Concept for Improving Airspace User Services,” 20th Digital Avionics Systems Conference, Daytona Beach, FL, Oct. 2001.

[9] K. Leiden, P. Kopardekar, and S. Green, “Controller Workload Analysis Methodology to Predict Increases in Airspace Capacity,” AIAA 2003-6808, AIAA Aviation Technology, Integration and Operations Conference, Denver, CO, Nov. 2003.

[10] T. Prevot, T. Callantine, J. Homola, P. Lee, J. Mercer, E. Palmer and N. Smith, “Air/Ground Simulation of Trajectory-Oriented Operations with Limited Delegation,” 7th USA-Europe ATM R&D Seminar, Barcelona, Spain, July 2007.

[11] D. McNally, E. Mueller, D. Thipphavong, R. Paielli, J. Cheng, C. Lee, S. Sahlman, and J. Walton, “A Near-Term Concept for Trajectory-Based Operations with Air/Ground Data Link Communication,” 27th International Congress of the Aeronautical Sciences, Nice, France, Sept. 2010.

[12] D. McNally, S. Engelland, R. Bach, W. Chan, C. Brasil, C. Gong, J. Frey and D. Vincent, “Operational Evaluation of the Direct-To Controller Tool,” Proceedings of the 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, NM, 2001.

[13] H. Swenson, T. Hoang, S. Engelland, D. Vincent, T. Sanders, B. Sanford and K. Heere, “Design and Operational Evaluation of the Traffic Management Advisor at the Fort Worth Air Route Traffic Control Center,” Proceeding of the 1st USA/Europe ATM R&D Seminar, Sacalay, France, 1997.

[14] D. McNally and C. Gong, “Concept and Laboratory Analysis of Trajectory-Based Automation for Separation Assurance,” Air Traffic Control Quarterly, Vol. 15(1), 2007.

[15] T. Farley, S. Landry, T. Hoang, M. Nickelson, K. Levin, D. Rowe and J. Welch, “Multi-Center Traffic Management Advisor: Operational Test Results,” AIAA 2005-7300, AIAA Aviation Technology, Integration and Operations Conference, Arlington, VA, Sept, 2005.

[16] B. Sullivan, R. Panda, and P. Soukup, “Implementing FANS in the NASA 747-400 Flight Simulator for Airspace Operational Research (Future Air Navigation System),” AIAA Flight Simulation Technologies Conference, Washington, D.C., 1996.

[17] E. Mueller, and S. Lozito, “Flight Deck Procedural Guidelines for Datalink Trajectory Negotiation,” AIAA 2008-8901, 26th Congress of the International Council of the Aeronautical Congress (ICAS), Anchorage, AK, Sept. 2008.

[18] T. Prevot, “Exploring the Many Perspectives of Distributed Air Traffic Management: The Multi Aircraft Control System MACS,” Proceedings of the International Conference on Human-Computer Interaction in Aeronautics, AAAI Press, Menlo Park, CA, Oct. 2002, pp. 149.

[19] D. McNally, H. Erzberger, R. Bach and W. Chan, “A Controller Tool for Transition Airspace,” AIAA-99-4298, AIAA Guidance, Navigation and Control Conference, Portland, OR, Aug. 1999.

[20] D. McNally, R. Bach and W. Chan, “Field Test Evaluation of the CTAS Conflict Prediction and Trial Planning Capability,” AIAA-98-4480, AIAA Guidance, Navigation and Control Conference, Boston, MA, Aug. 1998.

[21] H. Erzberger, T. Lauderdale and Y. Chu, “Automated Conflict Resolution, Arrival Management and Weather Avoidance for ATM,” 27th International Congress of the Aeronautical Sciences, Nice, France, Sept. 2010.

[22] H. Erzberger, “Transforming the NAS: The Next Generation Air Traffic Control System,” 24th International Congress of the Aeronautical Sciences, Yokohama, Japan, Sept. 2004.

[23] H. Erzberger, “Automated Conflict Resolution for Air Traffic Control,” 25th International Congress of the Aeronautical Sciences, Hamburg, Germany, Sept. 2006.

[24] E. Mueller, “Experimental Evaluation of an Integrated Datalink and Automation-Based Strategic Trajectory Concept,” AIAA 2007-7777, AIAA Aviation Technology, Integration and Operations Conference, Belfast, Northern Ireland, Sept. 2007.

[25] “Interoperability Requirements for ATS Applications using ARINC 622 Data Communications (FANS 1/A Interop Standard),” DO-258A, RTCA SC-189, Washington D.C., 2005.

[26] “PETAL Newsletter,” PETAL-to-LINK Programme, Eurocontrol, Vol. 4, No. 2, Apr. 2002.

[27] M. Mackintosh, S. Lozito, A. McGann and E. Logsdon, “Designing Procedures for Controller-Pilot Data Link Communication: Effects of Textual Data Link on Information Transfer,” AIAA 1999-01-5507, AIAA World Aviation Conference, Vol. 1, Washington D.C., Oct. 1999.

[28] W. Olson and N. Sarter, “Management by Consent in Human-Machine Systems: When and Why it Breaks Down,” Human Factors: The Journal of the Human Factors and Ergonomics Society, Vol. 43, No. 2, 2001, pp. 255-266, DOI: 10.1518/001872001775900904

[29] S. Hart and L. Staveland, “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,” Human Mental Workload, P.A. Hancock and N. Meshkati (eds.), North Holland Press, Amsterdam, The Netherlands, 1988, pp. 239-350.