Radio Communications Simulation and Aircrew Training

BRI-TR-130901

September, 2001

Alfred T. Lee, Ph.D., CPE

BETA RESEARCH INC

www.beta-research.com

18379 Main Blvd. Los Gatos, CA 95033 Tel: (408)353-2665 Fax: (408)353-6725

Radio Communications Simulation and Aircrew Training

Executive Summary

An increasingly congested air traffic environment emphasizes the importance of realistic radio communications simulation (RCS) in the evaluation and training of aircrews. The ability to realistically simulate air-ground radio communications in aircrew training environments is therefore paramount. However, a recent survey of U.S. airlines has revealed a remarkably low level of radio communications simulation fidelity. None of the airlines used trained air traffic controllers, but rely on instructor/evaluators (I/Es) to perform this task in addition to their instructional and evaluation duties. Most airlines do not provide the radio frequency chatter common in the operational environment nor do they provide additional staff to simulate non-ATC communications, such as dispatch and maintenance. I/Es are expected to provide most or all RCS in addition to their normal teaching and evaluation duties. The low level of RCS fidelity and the added burden of RCS on I/Es may affect the performance of both aircrews and instructors. A study was conducted to determine whether variations in RCS fidelity affect aircrew or instructor behavior in a simulated line-oriented evaluation. 12 current B747-400 aircrews were required to fly a flight scenario comparable to that they might receive in an actual training environment. The flight scenarios where flown in an FAA-certified flight simulator. Six crews were assigned to a low RCS fidelity condition comparable to the RCS level provided in training environments and six crews were assigned to a high RCS fidelity condition comparable to actual line operations. The two groups were matched for flying experience. Results indicate that crew planning behavior was reduced significantly in the high RCS fidelity condition. Crews also initiated ATC communications at a much higher frequency in the high RCS fidelity condition. I/E inter-rater reliability also increased significantly in the high RCS fidelity condition implying increased validity, as well as reliability, of instructor evaluations. Crew-system performance was also affected by RCS fidelity. High RCS fidelity led to substantially increased time for crews to execute a missed approach due to increased time spent on crew-ATC coordination and communication. These findings suggest that there is a marked influence of RCS fidelity on crew communications workload, on the reliability of I/E evaluations, and on crew and crew-ATC coordination processes. The absence of realistic RCS in current airline training facilities may be compromising crew evaluations. Recommendations for further research into the role of RCS fidelity in aircrew training environments are discussed.

Introduction

Radio communications plays a critical role in aviation safety in general and in airline operations in particular. This is due to the fact that civil aircraft, especially commercial aircraft, operate within a controlled airspace environment where radio communications between Air Traffic Control (ATC) and aircraft are needed in order to assure safe separation between aircraft. For commercial aircraft, which often operate in high-density traffic areas, clear and concise radio communications are essential. The worst airline disaster in history, the collision of two Boeing 747s at Tenerife in the Canary Islands, was due to poor radio communications. Communication problems have been implicated in a variety of airliner accidents since then (Kanki and Palmer, 1993). While critical clearance and traffic separation information is the primary focus of radio communications, it has other valuable uses. Information on weather conditions, traffic delays, as well as maintenance and dispatch information depend on air-ground radio communications.

Training in radio communications begins very early in a pilot's career. That training includes the operation of radio equipment as well as the use of proper communications procedures and phraseology. Before a pilot begins flying in commercial aircraft, he or she will have amassed several thousands of hours of experience during which radio communications will have become integrated into the full array of flying tasks. Thus, it is expected that pilots will have demonstrated proficiency in radio communications well before reaching airline operations.

It is perhaps this fact as well, as the secondary importance of radio communications in the pilot's task environment, that has resulted in the relatively low status of radio communications simulation (RCS) in the development of commercial airline flight simulators. Flight training simulators are critical to airline training and evaluation programs as neither is performed in the actual aircraft itself. A heavy investment in flight simulator development has culminated in a series of FAA flight training simulator standards which U.S. airlines are required to meet. While modern flight simulators are unparalleled in their delivery of realistic aircraft systems and flight task environment simulation, they lack the realism of actual radio communications environment that aircrews experience on a daily basis.

In a recent survey of U.S. airline instructor/evaluators (I/Es), the majority of respondents confirmed the low level of RCS fidelity[1] extant in flight training simulations (Burke-Cohen, Kendra, Volpe, Kanki, and Lee, 2000). Respondents noted that RCS for ATC radio communications is always simulated by the I/E in addition to his or her crew evaluation duties. The I/Es play the role of air traffic controller, dispatcher, maintenance officer, and any other role required of a simulation scenario at the same time they are observing, evaluating and instructing crews. The survey also revealed a very low number of training facilities that simulated radio communications

[1] Fidelity refers to the degree to which a simulation replicates either the physical elements of a real-world system (physical fidelity) or the degree to which the pilot perceives the system to be like that which exists in the real world (perceived fidelity).

traffic to other aircraft on the same frequency as the trainee's aircraft ("frequency chatter"). This is probably due to the high workload that instructors would incur if they attempted to simulate frequency chatter along with all the other duties they are required to perform. As monitoring of an assigned radio frequency is required of all pilots in real world operations, the lack of any frequency chatter in these simulations is patently unrealistic. The absence of simulator frequency chatter also eliminates the need for pilots to distinguish messages meant for his or her aircraft (ownship communications) from those meant for other aircraft on the same frequency. The inevitable result is a lowered pilot communications workload. Additionally, ATC or other ground sources are not cognizant of tasks that aircrews may be currently performing in the cockpit. As a consequence, these communications may serve as source of crew distraction in the real world, a distraction that crews must learn to manage. When I/Es simulate communications, however, they are much more likely to wait for crews to complete tasks before issuing a message thereby eliminating the distracting characteristic of these communications. The above shortcomings of current RCS methods and techniques appear to be understood by survey respondents in the Burke-Cohen, et al. study. A majority of the respondents believe that the low RCS fidelity of current training simulations results in an unrealistically low communications workload for aircrews.

While surveys can provide useful information on training practices and on I/E experiences and opinions, they cannot provide the necessary empirical evidence for any adverse effects that low RCS fidelity may or may not have on aircrew training and evaluation. Only a controlled experiment, one that would systematically vary RCS fidelity and objectively measure crew and I/E performance, can provide the necessary information.

Purpose

The present study was conducted in order to answer the question as to whether or not variations in RCS fidelity would impact crew performance or the evaluation of that performance by experienced I/Es. Specifically, the study sought to examine whether crews evaluated in a low RCS fidelity simulation environment would exhibit behavior that differ in significant and operationally meaningful way from those crews evaluated in a high RCS fidelity simulation environment. The high RCS environment defined here as being one which is comparable in physical or perceived fidelity with the radio communications experienced in line flight operations. In addition to the potential role of RCS fidelity in altering crew performance, this study also investigated how it might affect the evaluation behavior of the I/E. Of particular interest was the potential of low RCS fidelity for reducing the reliability or consistency of I/E crew ratings. This may occur if low RCS fidelity introduces anomalous crew behaviors that would not be expected under comparable line conditions. Such behaviors may lead to increased variability in crew evaluations among I/Es in their attempt to assess crew proficiency. Finally, the study also provided an opportunity to determine what elements of RCS appear to be important for producing a realistic RCS environment. Subsequent research could then focus efforts on assessing variations in RCS fidelity in a more cost-effective manner.

Method

Participants

A total of 24 current, B747-400 pilots composing 12 crews of two pilots each, participated in the study. All crews were paid for their participation. The crews were assigned to one of two groups of six crews each. Crews were assigned to each group based on flying experience in the aircraft type and series in order to form two separate groups of comparable experience. With two exceptions, the crews were composed of individuals from the same airline. In the low RCS fidelity group (LF), crews averaged 33.0 mos. of B747-400 flying experience while crews in the high RCS fidelity (HF) group averaged 38.0 months of B747-400 experience. The difference is not statistically significant.

Flight Simulator and RCS Fidelity

A level D, FAA-certified flight training simulator was used in this study. The flight simulator provided full motion and day visual scene simulation in accordance with FAA regulations for level D flight training simulators. The FAA certification meant that the simulator was of the same basic configuration and fidelity as other B747-400 simulators of this type found in the U.S. air carrier fleet. With the exception of the manipulations of RCS fidelity described below, the flight simulator configuration was identical for both low and high RCS fidelity groups.

RCS fidelity was varied between the two groups in the following manner. For the LF group, the I/E, a qualified B747-400 check airman, provided all radio communications required of the scenario from within the simulator. These included ATC, Automated Terminal Information System (ATIS), dispatch, and maintenance and the role of flight attendant. All of these roles where played in addition to the I/E's crew evaluation duties. No voice disguising system was employed. The I/E provided all RCS roles in the same voice. ATC instructions and ATIS were provided using the same scripting for both LF and HF groups in order that the same ATC and ATIS communications would be provided to both groups. No frequency chatter was provided to crews in the LF group. Additionally, while crews wore microphones, they did not wear headsets nor were they required to dial in radio frequency settings in order to receive or transmit radio communications. (The latter fidelity element was included in order to remove the perceptual-motor workload of radio system operations for LF crews).

In the HF group, the I/E provided no RCS though he did provide the flight attendant role. Individuals outside of the simulator provided all of the RCS in the scenario. Automated Terminal Information System (ATIS) transmissions were provided by a pre-recorded system that was activated when the pilot dialed in the appropriate ATIS frequency. A separate ATC station operated by trained air traffic controllers provided ATC to crews in the HF group. These controllers provided all ATC communications from ground, tower, center and approach/departure facilities. A voice disguiser was used to provide unique voice characteristics for communications from individual ATC facilities and sectors.

Frequency chatter was provided on all ATC frequencies throughout the simulated flight. All crewmembers wore headsets and operated radio communications equipment in the normal manner. The same I/E used in the LF group was used again in the HF group. Only the role-playing of a flight attendant was added to his normal I/E evaluation duties for the HF group.

For both groups, the I/E was required to provide crew performance ratings on a 5-point, Likert-type scale. Ratings were provided for five CRM skills (workload management, planning and preparation, crew communications, decision-making, ATC) and technical proficiency ratings for manual (TPM) and automation (TPA) skills. Ratings of crew performance were made for each of eleven separate flight phases.

Flight Scenario

The flight scenario used in this study was designed in accordance with published scenario guidelines for line-oriented flight simulations intended to evaluate aircrew CRM and technical proficiency skills (ATA, 1995). It consisted of a scheduled flight route between Los Angeles International Airport (LAX) and San Francisco International Airport (SFO), a flight duration normally of about 53 min. The aircraft system status, route, and fuel loading were consistent with an aircraft that had just arrived in LAX and was to be re-fueled for its flight to SFO. Weather conditions at LAX and SFO simulated those normally associated with a storm moving onshore from the northern Pacific. With such a storm, the most severe weather activity would occur in central and northern California. Crew briefings included forecasts of heavy rain as well as wind gusts and crosswinds in takeoff and landing. Light to moderate turbulence enroute and during the descent into SFO was to be expected.

In addition to weather-related problems at LAX, SFO and enroute, crews were also required to address an unanticipated nose gear malfunction during the approach into SFO. The nose gear malfunction alarm occurred approximately 30 sec. after the crew lowered the landing gear. This malfunction necessitated a missed approach at SFO in accordance with ATC instructions. ATC then ordered the simulated aircraft to hold at a nearby location in order to allow additional time for the reintegration of the aircraft in to the traffic flow. As the initial approach was on 28L at SFO (28R was not available at the time) the crews were now allowed to chose the longer 28R runway. As ATIS reports indicated fair braking action at SFO, the use of a longer runway would be more desirable. The use of the 28R runway would also permit the crews to use a CAT 3 autoland approach, a more desirable approach mode under the prevailing circumstances. These factors would strongly bias crews to change runways and this change, in turn, would require re-programming of the Flight Management System (FMS) as well as add ATC and crew communications, all of which were intended to increase crew workload during this stage of the flight.

Procedure

All crews were tested individually, one crew per day. Crews were not told the purpose of the study, only that they were to consider the simulator session as a line-oriented evaluation (LOE) and to behave accordingly. Crews were also told that a B747-400 check airman would accompany them as an observer and that the session would be videotaped. The crew pre-flight briefings were provided in accordance with typical line operations briefings that would be received from company dispatch. Crews were provided with the necessary charts, documents, and weather briefing materials appropriate for the flight route. They were instructed to fly in accordance with company procedures currently in effect for this aircraft type including fuel loading and any other safety minima. For both groups, the Captain (CA) served as pilot flying (PF) for the session and the First Officer (FO) served as pilot not flying (PNF). As is typical of airline operations, the PNF would perform all radio communications duties unless otherwise instructed by the CA.

Following the simulator session, each crew was individually de-briefed. At this time, crews were provided feedback on their performance by the check airman, and where appropriate, queried as to why certain decisions and actions were taken. Each crewmember then provided ratings and comments on simulator workload and ATC realism.

Results

The findings of this study are divided into four separate crew performance and instructor rating categories: CRM behavior, I/E ratings and rating reliability, crew-system performance measures, and crew ratings of workload and ATC realism.

Crew Resource Management Behavior

Crew Resource Management (CRM) behavior was assessed through analyses of communications that occurred between crewmembers and between a crewmember and others (ATC, dispatch, maintenance, flight attendants, and passenger announcements). Crew communications were decomposed into individual speech acts[2] for each crewmember from the time departure clearance was received in LAX to the time the aircraft's parking brake was set in SFO. Thus, communications in all flight phases throughout the flight scenario were included in the analyses.

The speech acts were assigned to four basic CRM categories based on their referential content: Situation awareness-aircraft (SitAC), situation awareness-airspace (SitAS), workload management (Workload), and planning/preparation (Planning). The SitAC category includes all communications that refer to any aspect of aircraft status or operation as well to the status of any object or individual within it (e.g., passengers).

[2] A speech act is a proposition that conveys the conceptual content of an utterance. Scripted communications such as ATC frequency chatter and ATIS information are not included in these data nor are checklist readouts and replies.

These communications would include aircraft speed, heading, altitude, fuel, control surface configuration (e.g., flaps) as well as references to FMS data and mode control panel displays. The SitAS category includes references to anything outside of the aircraft. These include air and ground traffic, weather, ATC clearances, and taxiway and runway observations. The Workload category includes communications regarding the distribution, deferment, or re-allocation of tasks among crewmembers. The Planning category includes all communications made with reference to the planning and preparation for future events. Included in this category are departure, takeoff and approach briefings, and contingency planning briefings. While these four categories are largely independent of each other, occasional overlap between categories may occur. For example, planning activities may also include how workload or tasks would be allocated between the two crewmembers in the case of some future event. In this case, the communications would be assigned to both Workload and Planning categories.

The results of the analyses of speech acts, averaged across flight phases, are shown in Figure 1. The mean percent speech act is defined as the average percent of total crew communications for that particular category for each RCS fidelity group. Analyses of Variance (ANOVA) of the mean percent speech acts revealed a significant effect of categories ($F=68.58$, $p<.01$). As can be seen in Figure 1, a large percentage of speech acts are devoted to SitAC and a much smaller percentage of speech acts to the SitAS category. This is to be expected as crews typically spend more time communicating about aircraft issues, then about airspace issues. For the HF group, both situation awareness categories increased slightly as did the percentage of communications in the Workload category. However, these differences are not statistically significant (Newman-Keuls, $p<.05$). A marked drop in Planning communications did occur in the HF group, however. This is reflected in the significant interaction effect between categories and RCS fidelity effects ($F=3.60$, $p<.05$). Planning communications dropped from an average of 34.8% of speech acts in the LF group to an average of 25.5% of speech acts in the HF group.

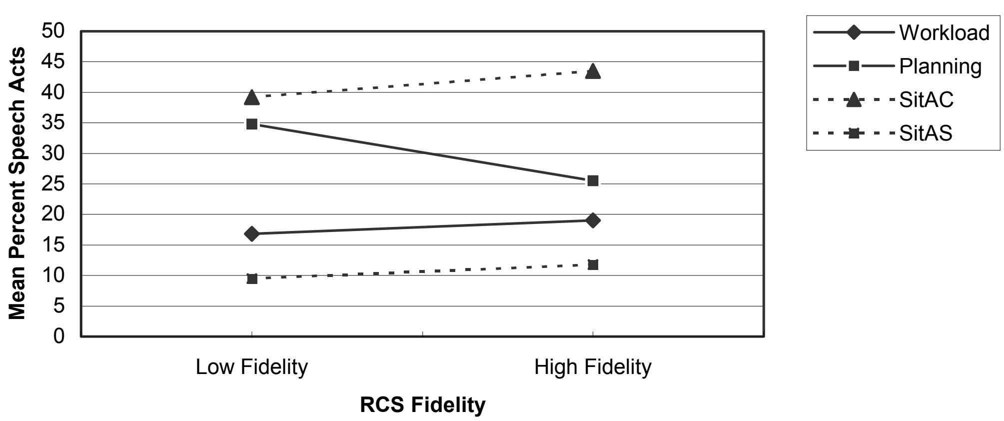

Figure 1. Mean speech acts for the Workload Management (Workload), Planning and Preparation (Planning), Situation Awareness-Aircraft (SitAC), and Situation Awareness –Airspace (SitAS) communication categories.

External Communications

Radio communications from crewmembers to those outside of the aircraft were also analyzed. Rather than analyzed as speech acts, however, these communications were analyzed as events. Thus, each time a crewmember initiated[3] a radio communication, it would count as a separate communications event, regardless of its duration. An event frequency measure was used rather than content analysis because of the highly proceduralized language characteristic of radio communications. These restrictions would not allow for much variation in content between groups. Further, the fact that the event is initiated by crewmember is probably more important than its content. Such radio communications consume time and effort and therefore are not normally done unless the perceived need for them is significant.

The overall frequency of these radio communications did not differ significantly between the RCS fidelity groups. HF crews totaled 131 such communications compared to an average 140 for LF crews. However, when the events are divided into ATC and non-ATC (i.e., dispatch, maintenance, flight attendant, passenger) communications, significant differences emerge between the two groups. As seen in Table 1, nearly twice as many ATC communications were initiated by crews in the HF group as were initiated in the LF group. The reliability of this finding is confirmed by a statistically significant difference in the interaction between the level of RCS fidelity and communication category ($F=14.33$, $p<.001$).

Table 1. Non-ATC and ATC Communications Initiated by Aircrews

	Non-ATC	ATC
Low RCS Fidelity	74 (12.3)	57(9.5)
High RCS Fidelity	37(6.2)	103(17.2)

Note: Means in parentheses

I/E Ratings

The reliability of I/E ratings of aircrew performance is particularly important to the aircrew training process, as these ratings are the primary means of assessing crew competency to perform their particular duties. Poor reliability of instructor ratings means not only that inconsistencies exist in the rating process, but may also imply that the ratings may lack validity as well, i.e., that they may not be measuring what they purport to measure. For this reason, I/E inter-rater reliability needs to be established before analyses of the ratings themselves can proceed with any confidence.

[3] Radio communications that were part of an ongoing communication transaction, such as ATC clearance readback or frequency change communications, are not included. Only those radio communications initiated voluntarily by a crewmember, without external prompting, are included.

I/E ratings in the present study were provided by one I/E during the simulation session and also by a second I/E after all sessions were completed. The second was also a qualified B747-400 check airman who rated crews using videotapes of the sessions. This I/E was unaware of the experimental manipulations and of the purpose of the study. Table 2 shows the inter-rater reliability for the two I/Es participating in this study as a function of RCS fidelity and rating category. The CRM rating category values are based on the I/E ratings averaged across the five separate CRM measures and all flight phases. The technical proficiency values are also based on the I/E ratings averaged across flight phases with separate ratings for technical proficiency in manual (TPM) and in automated skills (TPA).

Table 2. I/E Inter-rater Reliability for Crew CRM and Technical Proficiency in Manual Skills (TPM) and Automation Skills (TPA).

	CRM	**TPA**	**TPM**	**All**
Low RCS Fidelity	0.15	0.24	0.54	0.29
High RCS Fidelity	0.73	0.71	0.91	0.62

In every category, inter-rater reliability was substantially greater in the high RCS fidelity group with the largest increases found in the CRM categories. Ratings of technical proficiency in both manual and automation skills also increased in the high RCS fidelity group.

Because of the impact of the RCS fidelity manipulation on I/E inter-rater reliability, any interpretation of their ratings of crew performance as a function of RCS fidelity is necessarily confounded. As, the question of whether or not the crews in the two groups would rated be differently by the instructors cannot be answered.

Crew-System Performance

The influence of RCS fidelity on crew-system performance measures was also examined. These are direct measures of how crewmembers operated the aircraft or its component subsystems during the flight scenario. The first of these measures assessed the impact of RCS fidelity on the performance of crews in executing the missed approach procedure. The missed approach procedure was necessitated by the nose gear malfunction indication shortly following lowering of the landing gear. Missed approach performance was defined as the elapsed time from the onset of the audible nose gear malfunction alarm to the time at which go-around power was applied to the engines. This measure is affected by the distance from the runway at which the landing gear is lowered as less time would be available for the procedure to be initiated. To minimize this artifact, the two groups where made comparable by removing from the analysis a crew from the LF group which lowered its landing gear much earlier than other crews in its group and by removing from the analysis a crew from the HF group which lowered its landing gear much later than other crews in its group.

For the resultant data set, the missed approach performance time for the LF crews averaged 37.0 sec. while the HF crews averaged 82.8 sec. to execute the missed approach. The difference in time is statistically significant ($t=2.47$, $p<.05$). Thus, crews in the HF groups required substantially more time to execute the missed approach than those in the LF group. Analyses of communications transcripts suggests that the difference appears to be due to the additional ATC communications and coordination time required by crews in the HF group.

Two additional crew-system performance measures were also analyzed. These two measures examined how the crews operated the aircraft's advanced display and automation equipment. The Electronic Flight Instrument System (EFIS) displays information on the airspace environment as well as information relevant to an approach. Both crewmembers have such a display that can be set for one of four modes independently by each crewmember (map, approach, plan, and VOR). Examination of the use of these modes by individual crewmembers might reveal influences of ECS fidelity on, for example, situation awareness. However, no reliable differences were found in EFIS mode usage between groups, crew position, or their interaction.

An additional crew-system performance measure of interest was the use of the FMS/CDU (Control Display Unit) for programming activities. FMS programming activities were pronounced in this scenario due to the need to program an unpublished holding pattern and to program/verify an approach to a different runway. FMS/CDU inputs in the 15 min. period following the time at which the aircraft had reached its assigned missed approach altitude were examined for differences between the two RCS fidelity groups. An ANOVA of the input responses revealed significant differences between crewmember positions in the number of CDU inputs ($F=11.77$, $p<.005$). However, no reliable differences were found as a function of RCS fidelity or of RCS fidelity and its potential interaction with crew position. The latter finding is of particular interest as the CDU input task allocation between crewmembers could change as a function of the level of RCS fidelity if workload were affected. The fact that the FO had the largest input frequency was expected, as the FO was the PNF in the study and FMS programming is primarily a PNF task.

Crew Ratings

Following their simulator session, all crews were debriefed and were required to provide ratings in three categories using a 5-pt Likert-type scale. The scale required crewmembers to provide a rating comparing the simulator session they just experienced with 1) routine operational workload levels, 2) the simulator LOE ATC communications realism of their own training simulators, and 3) real-world ATC communication realism.

Figure 3 shows the average crew ratings for these categories for crews in each of the two RCS fidelity groups. When compared to routine operational workloads, no statistically reliable differences in ratings were found for the two groups. Both groups considered the level of workload for the routine flight operations elements of the simulator session to be comparable to what they routinely experience in actual operations.

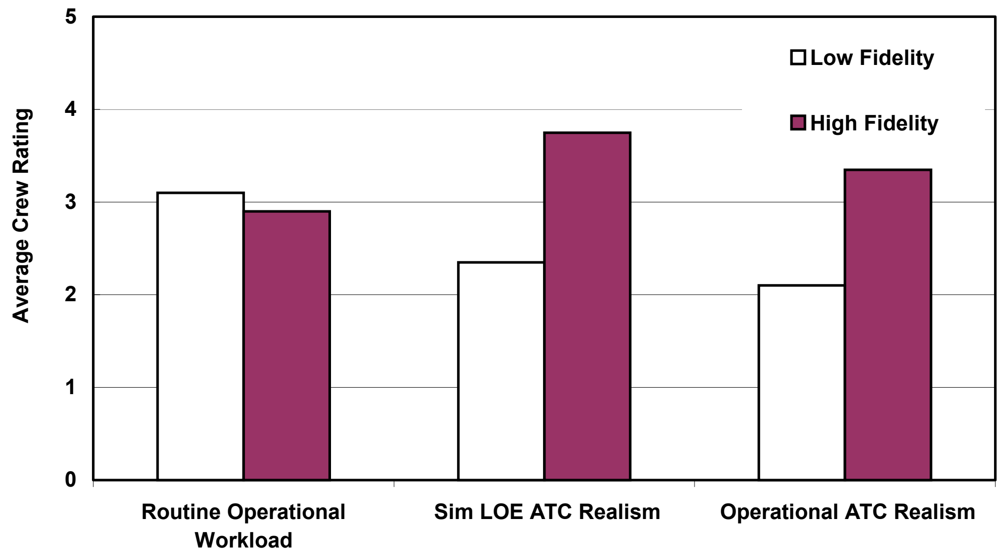

Figure 2. Crew ratings for simulator workload and realism.

RCS fidelity effects were found, however, in the last two rating categories. Crews in the LF group rated the ATC realism of their simulator sessions to be significantly lower than what they had experienced in simulators used for LOE in their own facilities (t=3.46, p<.05). Analyses of crew comments suggest that this was due to the lack of frequency chatter and the absence of a need to operate radio equipment in the low RCS fidelity condition. Crews in the high RCS fidelity commented that the presence of frequency chatter and use of different voices in each ATC facility in the high RCS fidelity condition substantially enhanced ATC realism compared to what they experienced in their own facilities. The final rating category asked crews to rate the ATC realism of the simulator session they had just completed with the level of ATC realism they experience in actual line operations. As can be seen in Figure 3, high RCS fidelity crews found the level of ATC realism in the present study to be comparable to that of line operations (Mean=3.35) while those in the LG group found ATC to realism to be markedly less (Mean=2.1). The difference is statistically significant (t=4.64, p<.001).

Discussion

The role of RCS fidelity was examined in this study by manipulating the level of RCS fidelity between two extremes (low vs. high) and then measuring the impact on crew and I/E performance. The manipulation attempted to provide a low level of RCS fidelity that might reasonably be expected to occur in an operational environment and a high level of RCS fidelity comparable to that which would occur in actual aircraft line operations.

The systematic manipulation of these two levels of RCS fidelity was successful in creating marked differences in perceived ATC realism by the participating aircrews. Thus, the variation in physical RCS fidelity resulted in a variation in perceived fidelity as evidenced by the crew ratings of perceived ATC realism in both groups when compared to actual line operations. The implication is that the differences in the level of RCS fidelity that were provided in the present study were sufficient to reveal any potential effects on crew or I/E performance.

Crew Workload

One of the prevailing hypotheses concerning low RCS fidelity in airline training simulations is that the training and evaluation environment provides an unrealistically low level of crew communications workload. The present study partially supports this hypothesis. Analyses of crew communications reveal significant declines in planning and preparation communications, generally considered of low importance relative to other crew tasks. As these communications are considered by some instructors[4] to occur excessively in current LOE simulations, their reductions under higher communications workload of the high RCS fidelity condition might be expected. The crews would replace excessive planning activities with more valued ones as workload increased.

Other communications, indeed, actually increased in frequency with high RCS fidelity. The frequency of ATC communications initiated by aircrews nearly doubled in the high RCS fidelity group when compared to the same communications in the low RCS fidelity group. In the low RCS fidelity environment, where the ATC role-playing I/E is only a few feet from the crew, crews could expect a level of awareness of their current situation that would not be available in the high RCS fidelity condition. This fact would tend to make additional communications superfluous. Furthermore, crews are much less likely to solicit information on external conditions (e.g., weather) from someone who is in their own cockpit (e.g., the I/E) and much more likely to request that information from an outside source.

While the effect on crew-initiated ATC communications was clear, RCS fidelity did not appear to have much of any effect on the aircrews' perception of routine workload levels in this study. Both RCS fidelity groups rated perceived workload levels during the routine segments of the scenario as comparable to those they experienced in actual line operations.

Crew-System Performance

Inducing crews to behave in the simulated aircraft in a manner comparable to that which might be expected in the actual aircraft is essential in order to achieve valid crew

[4] Capt. Noel Kane personal communication

evaluations. In other words, a flight simulator can only be an effective assessment device if there is no discernible difference between how a crew operates the real aircraft and how it operates a simulated one.

In this study, crew-system performance was markedly affected by RCS fidelity in some aspects, but not in others. Large differences were found in the time required for crews to execute the missed approach procedure. High RCS fidelity crews spent more than twice as much time executing the procedure as crews in the low RCS fidelity group. If we assume that the high RCS fidelity condition is more likely to elicit realistic crew behavior because it is comparable to real-world aircraft operations, than crews in the low RCS fidelity group are performing in a manner that is *not* comparable to how they would behave in the actual aircraft. This suggests that the validity of crew evaluations under low RCS fidelity is questionable. Moreover, it is particularly important that I/Es provide valid assessments of crew performance when the crew is required to perform under unusual or safety-critical situations. The finding of large performance differences in executing a missed approach procedure in this study is therefore particularly significant.

While missed approach procedure performance was significantly affected by RCS fidelity variations, other crew-systems measures were not. Operation of advanced information display and automation equipment as represented by measurements of the use of EFIS displays and FMS/CDU programming activities did not reveal any differences between the two RCS fidelity groups. The implication of these findings suggests that RCS fidelity effects are localized to those areas that directly involve crew-ATC communication and coordination. They do not appear to generalize to overall crew workload or workload management activities or to aircraft or airspace situation awareness. Further, the perceptual-motor effort of radio equipment operation present in the high RCS fidelity condition did not appear to affect other perceptual-motor activities such as FMS/CDU programming. Analyses of de-briefing comments by aircrews did show that radio equipment operation affected the perceived realism of the simulation.

I/E Ratings

An essential element of the crew evaluation process in airline training regimens is the I/E. This individual has the authority to determine whether or not a pilot has the knowledge and skill necessary to safely pilot an airliner in today's complex airspace system. For this reason, the I/E ratings of crews must have a high degree of reliability and validity.

The present study revealed that the reliability of I/E crew ratings was markedly reduced for crews evaluated under low RCS fidelity conditions. One possible explanation for this finding is that low fidelity simulations are likely to elicit anomalous and highly varied behaviors as crews attempt to deal with the poorer simulator realism. When crews produce these kinds of behaviors, it becomes more difficult for I/Es to fit them to any existing standard or evaluation framework. The result is less reliable or consistent I/E ratings.

Regardless of the reason for the lowered rater reliability, the very existence of it and the fact that it occurred for both CRM and technical proficiency ratings has important implications for simulator-based crew evaluations in current air carrier training environments.

Conclusions

Several conclusions can be drawn from the present study concerning the role of RCS fidelity in simulator-base crew evaluations. The first is that RCS fidelity does affect crew performance in an operationally meaningful and significant manner. The effects appear to be primarily in areas of crew communications workload, in the coordination activities between crewmembers and between a crewmember and ATC. The latter will determine how crews will perform when under situations where ATC communications and collaboration are needed. These would include runway changes, missed approaches, runway incursions, weather hazard avoidance, and airport diversions.

RCS fidelity did not appear to affect overall crew workload or how the crews managed their workload over the course of the simulated flight. Nor did it affect crew situation awareness of aircraft systems or of the airspace environment. While crew-systems performance in executing a missed approach procedure was affected, operation of key aircraft subsystems was not.

Generalizing these findings to the operational training environment must be done with caution, however. First, the study examined RCS fidelity in line-oriented evaluation simulation and did not address how fidelity would affect skill acquisition, skill maintenance, or instructional efficacy. Secondly, the study was conducted in a research setting where crew and I/E performance would not have been under the same level of scrutiny as would occur in real training environments nor would the consequences of poor performance be as important to the professional careers of the participants. While observations of their performance suggest that these crews behaved in a professional manner, it is unlikely that the same level of motivation to perform existed in this study as would exist in operational training. Finally, it was not possible to isolate the influence of "frequency chatter" on crew behavior due to the limitations placed on the study by crew availability for the particular aircraft simulated. For all of the reasons, it is recommended that a subsequent study examining RCS fidelity effects be conducted in actual training environment. This would allow replication of the present study, isolation of the effects of frequency chatter as well as permitting a more robust examination of I/E reliability than was possible in the present study.

References

Author. *ATA Working Group Paper, LOFT/LOE Design*. Washington, D.C.: Air Transport Association, 1995.

Burke-Cohen, J., Kendra, A.J., Volpe, J.A., Kanki, B.G., Lee, A.T. *Realistic Radio Communications in Pilot Simulator Training*. DOT/FAA/AR-00/62. Washington, D.C.: Office of Aviation Research, December, 2000.

Kanki, B.G. and Palmer, M.T. Communication and Crew Resource Management. *In Cockpit Resource Management* (E.L. Weiner, B.G. Knaki, R.L. Helmreich, Eds). New York, NY: Academic Press, 1993.