AIR TRAFFIC CONTROL COMMUNICATIONS SIMULATION AND AIRCREW TRAINING

Alfred T. Lee, Ph.D.
Beta Research, Inc.
Los Gatos, CA USA

Abstract

Air Traffic Control (ATC) communications are an essential element of all airline flight operations. The provision of ATC communications in flight simulation technology would therefore seem to be a foregone requirement. Yet despite decades of rapid technical developments in flight simulators, these devices are unable to provide the necessary level of ATC communications simulation fidelity required for realistic line oriented simulation (LOS) training. Surveys of airline training organization indicate that only a minimal level of ATC communications simulation is provided and that it is provided manually by instructor/evaluators (I/Es), playing the role of air traffic controllers. In most cases, no background frequency chatter is provided. One of the reasons for the low level of ATC fidelity in airline training is the result of limitations in technology, but an additional reason is the lack of understanding of the benefits that a high level of ATC fidelity might bring to LOS training. In order to identify the contribution of high ATC fidelity to this type of training, three studies were conducted. The first study assessed the effects of ATC communications simulation fidelity on crew performance in a large air transport research simulator. The performance of current, type-qualified airline crews were evaluated with low or high fidelity ATC simulation in an LOS scenario. Significant differences were found in aircrew communications and performance. Differences were particularly evident in the content of intracrew communications and in the communications to sources outside the cockpit, including ATC. A second study, conducted at a major airline training center, evaluated ATC fidelity under actual crew evaluations for Captain upgrade training. This study revealed that high ATC fidelity also changed the rate of intracrew information exchange as well as the pattern of communications to external cockpit resources. A third study examined the effects of one aspect of ATC simulation fidelity, frequency chatter, on crew behavior. Mental workload ratings increased significantly with the presence of simulated frequency chatter, but no effects were found on crew performance. The results of these studies reveal that high ATC communications fidelity in LOS training has a significant impact on aircrew communications workload management and the use ATC as a resource. High ATC communications simulation fidelity may also affect crew mental workload and a crew's use of ATC resources. Implications for crew training and evaluation are discussed.

Introduction

Current airline aircrew flight training relies heavily on flight simulators for the training and evaluation of aircrews. At it's most advanced level of application, flight simulators are used to recreate the entire spectrum of flight operations. This full mission simulation capability or line-oriented simulation (LOS) is now common to airline training around the world. Recreating real life flight operations is complex and expensive. Simulators must provide a high degree of physical as well as perceived (or psychological) fidelity with the actual aircraft and its operating environment. The heavy reliance on these simulators for training and evaluation requires that the simulations have a high enough fidelity to elicit crew behaviors comparable to those that might be expected under actual aircraft operating conditions.

The technical complexity and cost associated with high fidelity flight simulation means that the goal of full fidelity may not always be met. One component of flight simulation, ATC communications, is far less developed then many other areas of flight simulation technology. A recent survey of U.S. air carrier instructor/evaluators (I/Es) indicates that virtually all ATC ownship[1] communications simulations are provided manually by the I/E (Ref. 1)). I/Es are not trained in ATC procedures and have only their flying experience to aid them in a task normally carried out by a trained professional. Significant differences in ATC ownship communications content, presentation rate, and communications procedures from those that crews experience in line operations are therefore inevitable in airline simulators. Frequency chatter, communications that are directed to other aircraft over a monitored radio frequency, are seldom provided. Part of the reason for the slow

[1] Ownship communications are those directed by ATC to the simulated aircraft being flown by the trainees

advancement of ATC communications simulation is the technical limitations inherent in current voice recognition and synthetic speech (Ref. 2). An additional impediment is the absence of evidence that providing a high level of ATC communications simulation fidelity is worth the investment. There is little incentive to improve the existing flight simulation unless some clear benefits can be demonstrated in crew training and evaluation, I/E performance, or crew or I/E acceptance of the value of the technology in enhancing training. Finally, there is no current regulatory requirement that ATC communications simulation be provided in airline simulators. While there are extensive requirements for visual scene simulation, motion cueing and other aspects of simulation, there is no requlatory impetus for providing ATC communications simulation.

Study 1

The present study was conducted in order to answer the question as to whether or not variations in ATC fidelity would impact crew performance. Specifically, the study sought to examine whether crews evaluated in a low ATC fidelity simulation environment would exhibit behavior that differs in a significant and operationally meaningful way from those crews evaluated in a high ATC fidelity simulation environment. The high ATC environment defined here as being one which is comparable in physical or perceived fidelity with the radio communications experienced in line flight operations. The study also provided an opportunity to determine what elements of ATC appear to be important for producing a realistic ATC environment. Subsequent research could then focus efforts on assessing variations in ATC 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 in a matched group design based on flying experience in this aircraft type and series. With two exceptions, the crews were composed of individuals from the same airline. In the low ATC fidelity group (LF), crews averaged 33.0 mos. of type-specific flying experience while crews in the high ATC fidelity (HF) group averaged 38.0 months of experience. The difference is not statistically significant.

Apparatus

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. With the exception of the manipulations of ATC fidelity described below, the flight simulator configuration was identical for both low and high ATC fidelity groups. 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 (Ref. 3).

ATC fidelity was varied between the two groups in the following manner. For the LF group, the I/E, a qualified 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 ATC 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.

In the HF group, the I/E provided no ATC, though he did provide the flight attendant role. Individuals outside of the simulator provided all of the ATC in the scenario. Automated Terminal Information System (ATIS) transmissions were provided by a pre-recorded system that was activated when the pilot entered 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 more 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.

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 if it were a line-oriented evaluation (LOE) and to behave accordingly. Crews were also told that a 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 performed 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 three separate crew performance categories: CRM behavior, crew-system performance measures, and crew ratings of workload and ATC realism.

CRM Behavior

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.

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). These communications would include aircraft speed, heading, altitude, fuel, control surface configuration (e.g., flaps) as well 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 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 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 that ATC fidelity group. Analyses of variance (ANOVA) of the mean percent speech acts revealed, as expected, a significant effect of categories ($F=68.58$, $p<.01$). A large percentage of speech acts were devoted to SitAC and a much smaller percentage of speech acts to the SitAS category. 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 ATC 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.

[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.

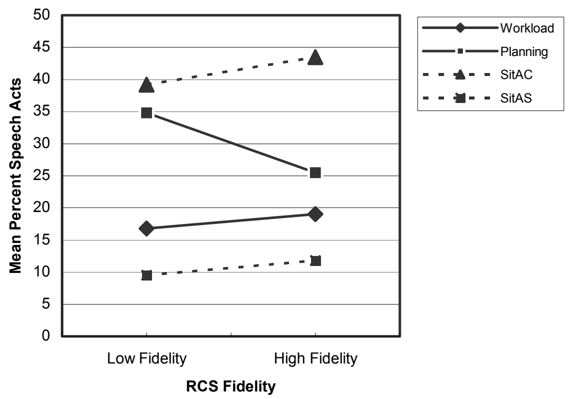

Figure 1. Mean speech acts for the Workload Management (Workload), Planning and Preparation (Planning), Situation Awareness-Aircraft (SitAC), and Situation Awareness –Airspace (SitAS) communication categories.

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 communication event, regardless of its duration. An event frequency measure was used rather than content analysis because of the very restricted phraseology of ATC 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 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 ATC 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 (e.g., dispatch, maintenance) 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. A statistically significant difference (χ^2=25.3, p<.01).

[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.

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

	Non-ATC	ATC
Low ATC Fidelity	74	57
High ATC Fidelity	37	103

Crew-System Performance

The influence of ATC 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 ATC 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 the time available for the procedure to be initiated depends on how far the aircraft is from the runway threshold. 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 ATC 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 ATC 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). No reliable differences were found as a function of ATC fidelity or of ATC 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 ATC 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 2 shows the average crew ratings for these categories for crews in each of the two ATC 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.

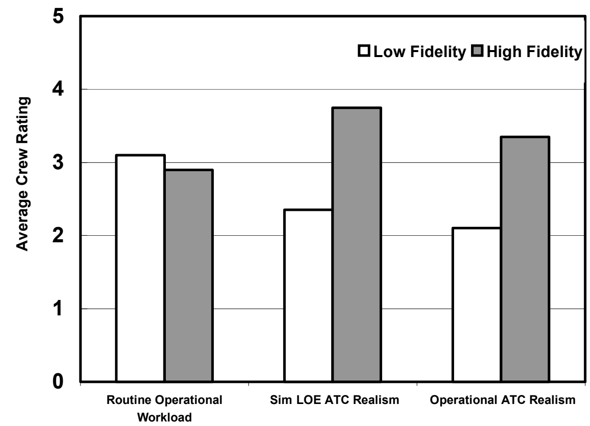

Figure 2. Crew ratings for simulator workload and realism.

ATC 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 ATC fidelity condition. Crews in the high ATC fidelity commented that the presence of frequency chatter and use of different voices in each ATC facility in the high ATC 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 2, high ATC 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 LF 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 ATC fidelity was examined in this study by manipulating the level of ATC 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 ATC fidelity that might reasonably be expected to occur in an operational environment and a high level of ATC fidelity comparable to that which would occur in actual aircraft line operations.

The systematic manipulation of these two levels of ATC fidelity was successful in creating marked differences in perceived ATC realism by the participating aircrews. Thus, the variation in physical ATC 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 ATC fidelity 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 ATC 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 ATC 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 ATC fidelity. The frequency of ATC communications initiated by aircrews nearly doubled in the high ATC fidelity group when compared to the same communications in the low ATC fidelity group.

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

[4] Capt. Neal Kane personal communication

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 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 ATC 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 ATC fidelity crews spent more than twice as much time executing the procedure as crews in the low ATC fidelity group. If we assume that the high ATC fidelity condition is more likely to elicit realistic crew behavior because it is comparable to real-world aircraft operations, than crews in the low ATC 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 ATC 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 ATC 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 ATC fidelity groups. The implication of these findings suggests that ATC 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.

Study 2

The above study data were collected in a research environment where the pressures on aircrew performance may not have been comparable to actual airline training situations. A second study was conducted to replicate this study in an actual airline training environment where crew and instructor motivation to perform would be much higher and where a larger number of I/Es could be included in the study. Additionally, if significant changes in crew performance can be demonstrated in an actual airline training environment there will be an additional impetus to develop and deploy improved ATC communications simulation technology in flight training simulators. The working hypothesis for the present study, as in the previous study, is that high ATC communications fidelity serves to provide the necessary secondary task loading for the crew that is normally provided in actual line operations. This, in turn, might affect how other tasks, including tasks other than frequency monitoring, are performed by crews.

Method

Participants

A total of 10 two-person crews participated in the study. Five crews were assigned to each of two groups in a matched group design. Groups were matched for crew time in aircraft type. Time in type averaged 3,500 hrs for those in the low fidelity (LF) group and 3,800 hrs. for the high fidelity (HF) group, a difference which is not statistically significant. Each participant was enrolled in the Captain upgrade program at the participating airline training organization. The program culminated in a line-oriented flight training (LOFT) session in the flight simulator. Only one of the crewmembers, the one serving in the Captain's role, was evaluated by an I/E during each LOFT, though the performance of both crewmembers was of interest in this study. To avoid any order effects, a crew served only once in the study. The six I/Es in the study were all senior check airman for the airline, averaging about 14 yrs. in instruction for the simulated aircraft.

Apparatus

A Level C, FAA-certified flight simulator was used in this study. The same flight simulator was used throughout the study to avoid the introduction of any simulator differences between conditions. For the LF crews, an I/E provided all ATC ownship communications and all other non-ATC communications (e.g., dispatch, flight attendant) to the simulated aircraft. These tasks were in addition to his normal duties of operating the simulator and observing/evaluating crew performance. For the HF crews, a professional air traffic controller provided all ATC ownship communications and frequency chatter. The frequency chatter included localized[5] airspace information including calls to other company traffic. In the high HF group, non-ATC communications were provided by the I/E who, as in the LF group, also operated the simulator and evaluated crew performance. The LOFT scenarios used in the simulator were part of the normal Captain upgrade LOFT training system and were approved for this use by the FAA. These scenarios were not altered in any way for the purposes of this study. Each scenario lasted approximately 90 min and required a diversion to an alternate airport. Scenarios typically included a system malfunction in addition to the diversion.

[5] Information about the airspace which is unique to the locale, e.g., specific departure route clearances

Procedure

All aircrews in this study had bid for, and served in, the six week Captain upgrade training program of the airline. The LOFT session served as the last event in the program and the last potential obstacle to promotion. All LOFT scenarios were conducted as they normally would have been by the I/Es. There was no attempt to alter the way in which I/Es conducted the LOFTs, with the exception of those in the HF condition where I/Es no longer provided ATC ownship communications. I/Es were briefed on the general goals of the study regarding ATC communications simulation, but were not aware of any of the specific hypotheses being tested. Aircrews were not aware of the purpose of the study and were assigned to the two groups based only on their aircraft time and availability. Following completion of the simulator sessions, all aircrews and I/Es completed individual questionnaires on the role of ATC communications, the fidelity of the simulation they had just completed, and workload ratings.

Results

In order to evaluate the effects of enhanced ATC communications simulation fidelity on crew and instructor behavior, three areas of crew performance were analyzed: Intracrew communications[6] (communications between pilots), pilot-initiated communications to sources outside the cockpit, and subjective ratings by crews and I/Es. These measurements were selected as being those most likely to be sensitive to variations in communications workload[7].

[6] Checklist tasks, briefings, scheduled or other procedural communications are not included these analyses

[7] Captain trainees were scored by I/Es on a pass/fail system, all trainees in this study passed the LOFT evaluation

Intracrew Communications

Communications between the PF and the PNF were analyzed to determine whether enhanced ATC communications fidelity would have an effect on them. One measure of these intracrew communications is the rate of exchange of communications between crewmembers. An index of this is the number of turns in communications per unit time. In any communication, each separate sender and recipient communication is considered a "turn". The number of these turns per unit time is the *turn rate* of the communications between crewmembers. High turn rates imply high rates of information exchange between crewmembers. Low rates imply low rates of exchange or that one or both members are engaged in extensive, explanatory communications as might occur during pre-departure briefings. In any case, a significant difference in turn between crews in LF and those in the HF condition would suggest that information exchange between crewmembers is being affected by ATC simulation fidelity. Analyses of intracrew communications in the present study found a mean turn rate for crews in the LF condition of 2.79 while the turn rate for those in the HF condition was 2.36. The difference is statistically significant (Mann-Whitney U =11, p<.01).

Pilot-Initiated Extra Cockpit Communications

Pilot-Initiated Extra Cockpit Communications (PECs) were also analyzed to determine whether or not ATC fidelity would affect this type of pilot task. PECs are those communications which are initiated voluntarily by a pilot without prompting by an external source. These include ATC contacts as well non-ATC communications to company (e.g. dispatch) or cabin (e.g., flight attendant) sources. Figure 3 shows the number of PECs for communications in each of the two groups. While average number of total PECs (31) was approximately the same for crews in both groups, the distribution of the PECs differed for the two groups. While there were no significant differences in the percentage of non-ATC and ATC PECS for the LF group (49% vs. 51%), significant differences were found in the average percentage of PECs devoted to non-ATC communications (43%) when compared to ATC communications (57%) in the HF group (t=2.31, p<.01).

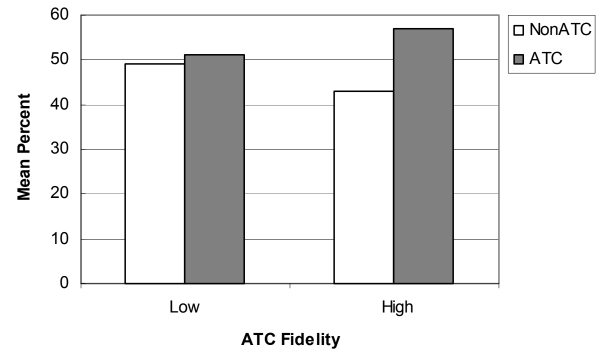

Figure 3. Mean percent Pilot-Initiated Communications as a function of ATC communications fidelity.

A further analysis of the PECs in the HF condition revealed that, while the percentage of company communications remained essentially the same, a reduction in the percentage of cabin communications and an increase in ATC communications occurred for crews in this group (see Figure 4).

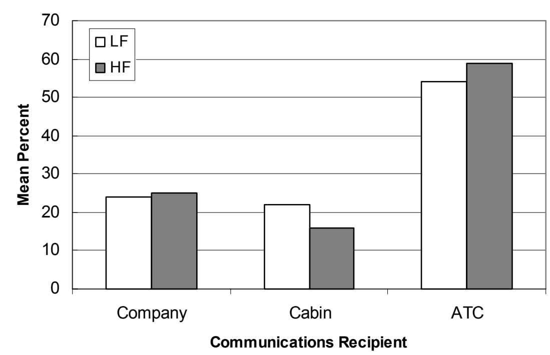

Figure 4. Pilot-initiated external cockpit communications in the high ATC fidelity condition.

Communications Procedural Errors

Pilot communications with ATC were also examined for communications procedural errors in this study. These include readback errors, use of the wrong radio frequency, use of incorrect callsign number, wrong or no facility name when contacting an ATC facility, missed ATC calls, and failure to make a facility contact. Call sign confusions, where the pilot contacted ATC in the belief that the message was for his aircraft, were also included in the analyses. Call sign confusions could only occur in the HF condition as only during this condition did crews receive communications intended for other aircraft (i.e., frequency chatter). Figure 5 shows the mean percent communications procedural errors that occurred in each of the two groups. The percent of communication procedural errors for each crew is based on the number of ATC ownship communications received by the crew. As with PECs, ownship communication events are defined here as those communications that are initiated by ATC and do not include those which are made in response to pilot transmissions. Although significant differences occur between the LF and HF crews in the percentage of communications procedural errors, the difference disappears almost entirely when call sign confusions are excluded from the analyses. With the exception of the call sign confusions associated with frequency chatter in the HF condition, no significant differences were found between the two groups in the number of communications procedural errors.

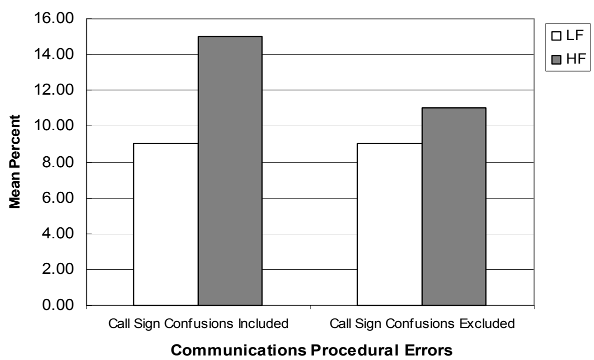

Figure 5. Communications procedural errors as a function of ATC fidelity.

Crew and Instructor Ratings

Following the completion of their LOFT session, each crewmember was requested to complete a questionnaire concerning ATC communications simulation. Figure 6 shows the average ratings received for two questions regarding ATC fidelity. The first question asked crewmembers to compare the workload of routine events (e.g., takeoff) in the simulator to those they experienced in the actual operational environment. While there is an increase in rated routine workload for those in the HF condition when compared the LF condition, the difference was not statistically significant. When asked to compare their simulator experience to those of prior LOFTs, crews in the HF condition rated their simulator experience as significantly greater in operational realism than crews in the LF condition ($t=3.0$, $p<.01$).

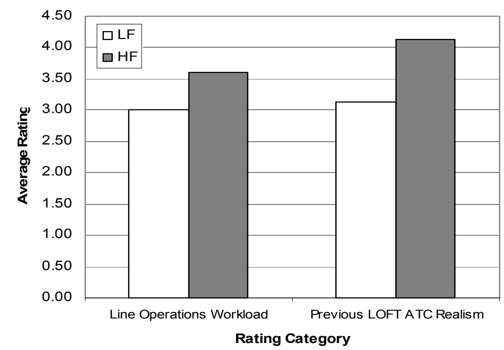

Figure 6. Crew ratings of ATC fidelity and realism
Instructor ratings were also collected in the study.

The six I/Es in this study provided feedback on the workload impact as well as the importance of ATC communications simulation in LOFT sessions. The six I/Es rated the effects of providing ATC ownship communications as being low to moderate, a rating in line with previous studies on the impact of ATC simulation task on I/E workload (Ref. 1). When asked to rate the importance of realistic ATC communications simulation in Captain upgrade training, instructor ratings averaged from very important to essential.

Discussion

The present study was conducted to investigate the effects, if any, of high fidelity ATC communications simulation on aircrews undergoing LOFT evaluation in an actual airline training environment. While Study 1 was conducted in a research environment which allowed for considerable control over simulation variables, the present study provided the opportunity for the assessment of ATC fidelity effects on crew behavior under the more motivating environment of a Captain upgrade LOFT evaluation. The study also allowed feedback from a larger sample of I/Es than was possible in a research environment.

The effects on communications workload management were most evident in the present study. The reduced rate of communications exchange between the two pilots and the change in pilot external communications which resulted from the higher ATC fidelity supports the hypothesis that high ATC fidelity produces crew communications behavior which is significantly different than that currently provided by low ATC fidelity in airline training environments. These differences are likely due to the overall increase in communications workload demanded by the high ATC fidelity condition, a workload level more like that of line operations. Higher ATC fidelity also tended to elicit more effective crew use of ATC as a resource. More realistic ATC communications simulation did not, however, affect the overall rate of communications procedural errors committed by aircrews in this study - with the exception of call sign confusions. Call sign confusions occurred as a direct result of the frequency chatter provided in the high ATC fidelity condition. Confusing one's own call sign with that of another aircraft on a radio frequency is not uncommon in airline flight operations. Call sign confusion, typically the result of similarities between ownship and other aircraft call signs on the same frequency, reflect the lack of full attention paid to ATC communications by some aircrews. This crew performance deficiency can only be observed in LOS when the frequency chatter component of high fidelity ATC is available. It would be important for the I/E to be able to identify this deficiency and correct it in training in order to reduce its occurrence in actual operations (Ref. 4, 5).

The increase in I/E workload resulting from the need to provide even a low level of ATC fidelity is evident from the ratings and comments received by I/Es in this study. Clearly, there is an impact on I/E workload, but it is not clear whether or how this affects the ability of the I/E to evaluate crew performance. It is possible that the low level of ATC fidelity that is provided is such as to not affect I/E crew training or assessment, though any impact on I/E workload which is not relevant to the task of training and evaluation is inherently undesirable.

Study 3

In the above studies, ATC communications fidelity was defined as low when an I/E provided ATC ownship communications and no frequency chatter was simulated. When a professional controller provided all ownship communications and frequency chatter was simulated, ATC fidelity was considered high. The low ATC fidelity condition in these studies was intended to simulate the level of ATC fidelity currently provided by airline training organizations. The high fidelity ATC condition was designed to approximate real-world ATC communications, a belief supported by crew comments and ratings in both studies. However, neither of these two studies addressed the independent contributions of either ownship communications or frequency chatter simulation. Identifying the relative contributions of each to crew training and evaluation is important if a cost-effective ATC simulation is to be achieved. The present study was conducted in order to isolate the influence, if any, of one of the two key elements of ATC communications simulation – frequency chatter. The choice of this element for evaluation was based on the likelihood that, for technical and cost reasons, it is most likely to be the first element of ATC communications simulation implemented in airline simulators. While the provision of frequency chatter in flight simulators is not without technical challenges, these are more likely to be overcome within the near future than those facing ownship communications simulations (see Ref. 2). Secondly, frequency chatter simulations are amenable to alternative, lower cost implementations which make it a much more attractive to training organization in meeting ATC communications simulation needs. For example, some training organizations have experimented with frequency chatter simulation using pre-recorded, real-world radio frequency transmissions. Finally, the absence of frequency chatter in flight simulations, particularly line-oriented simulations, is an obvious shortcoming in simulator design given the high density communications traffic typical of air carrier operations. Both aircrews and instructors are aware of the lack of ATC realism when frequency chatter simulation is absent. The reasons cited above makes the provision of frequency chatter simulation an attractive candidate for any proposed improvements to ATC communications realism in flight simulators.

This study was designed to assess the effects, if any, of realistic frequency chatter simulation on aircrew performance in a line-oriented simulation. Of particular interest was the effect frequency chatter simulation might have on crew workload since frequency monitoring by aircrews is a common secondary task in real-world operations. The presence of realistic frequency chatter simulation may increase workload which, in turn, might affect other aspects of crew performance.

Method

Participants

Twelve instrument-rated pilots holding commercial licenses participated in the study. Pilot ages ranged from 21 to 51 yrs. (Mdn=26 yrs.). Eleven of the twelve pilots were Certified Flight Instructors. Pilot flying experience ranged from a total of 265 to 14,000 hrs (Mdn=1,100 hrs.) with total instrument pilot-in-command hours ranging from 15 to 2,174 hrs (Mdn=72 hrs.). The pilots were randomly assigned to one of six two-person crews with the constraint that age and experience differences between the pilot-flying (PF) and the pilot-not-flying (PF) were minimized within each crew. In order to allow for a comparison between PF and PNF pilots, assignments to these positions were made so that the average age and experience of pilot crew position were similar. Participants were instructed to perform crew functions, including crew briefings and checklists, as they would in actual flight operations.

Apparatus

An FAA-Approved, Level 1 Flight Training Device (FTD) was used in the study[8]. The FTD simulated a light, twin-engine, piston-powered aircraft comparable in performance to a Piper Seminole. In order to provide realistic, radio frequency chatter (FC), recordings were obtained of actual frequency chatter corresponding to the planned scenario routes. The recordings were then edited to remove communications which did not conform to weather conditions within the flight simulation scenario. Additional editing of the recordings was performed in order that the number of transmissions per minute[9] was comparable to that of high density terminal areas. All ownship communications (communications directly to the simulated aircraft) were provided by a professional air traffic controller. While FC was provided only on airport approach and departure frequencies in this study, ATC ownship communications were provided on all communications frequencies including Clearance Delivery.

Procedure

Prior to the beginning of each simulator test session, pilots were provided with route clearance, weather, fuel available and other information pertinent to the simulated flight scenario. Each crew was also provided with a 30-min. simulator familiarization period to acquaint them with simulated aircraft operations, controls and displays. This familiarization period included a takeoff and an ILS approach. The test session consisted of two legs between three local airports, each leg was approximately 35 min. in duration. Both legs were flown in Instrument Meteorological Conditions (IMC) with occasional mist and rain showers. (There were no aircraft system malfunctions or unusual weather phenomenon, such as icing, in the simulation scenario). One of the two legs contained frequency chatter on airport departure and arrival frequencies in addition to the ownship communications provided by the controller. The other leg provided only ownship communications and no frequency chatter. Order of the legs was counterbalanced across crews so that half of the crews flew the leg containing the frequency chatter first while the remaining crews flew the frequency chatter leg second. The counterbalancing was intended to minimize order effects. Following the completion of each leg, crews were required to complete the NASA-TLX workload rating scale (Ref. 6). The scale contains six individual rating categories with an accompanying description of what each category is intended to measure. Each pilot completed the rating scales independently. Following the completion of the test session, each pilot was required to complete a questionnaire which solicited ratings, comments and suggestion of the value of ATC communications simulation for pilot training.

Results

Data from this study were organized into three separate categories for analyses: 1) pilot workload rating data from the NASA-TLX workload index, 2) communications and instrument procedure errors committed by pilots during the simulation, and 3) pilot responses to the post-study questionnaire on the value and realism of ATC communications simulation.

NASA-TLX Workload Ratings

The NASA-TLX is a workload index consisting of six separate rating scales. Each scale is accompanied by a description of the workload category to which it refers. The six rating scales are: Mental Demand, Physical Demand, Temporal Demand, Effort (physical and mental), Performance and Frustration. For the purpose of analyses, the original scale was converted to a 10-pt scale.

Mean workload ratings are shown in Figure 7 as a function of frequency chatter condition collapsed across pilot position (PF, PNF). While differences appear in every rating category, the only statistically reliable differences were found in the Mental Demand ($t=2.12$, $p<.057$) and in the Effort ($t=2.77$, $p<.018$) categories. In both categories, the presence of frequency chatter increased the pilots' required mental demand and effort. The increase in rated workload for the Effort command is likely due to increases in mental effort from the frequency monitoring task as actual radio communications operations (e.g., changing a radio frequency) are the same in both FC and No-FC conditions.

[8] An AST-300 manufactured by Aviation Simulation Technologies

[9] Combined air-to-ground and ground-to-air transmissions averaged approximately 11 per minute. A 10-second period was added within each minute where no frequency chatter occurred to allow for ownship communications.

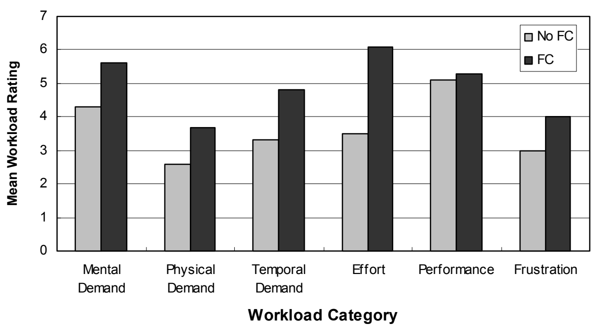

Figure 7. Mean workload ratings as a function of Frequency Chatter (FC) and No Frequency Chatter (No FC) conditions. (n=12 pilots).

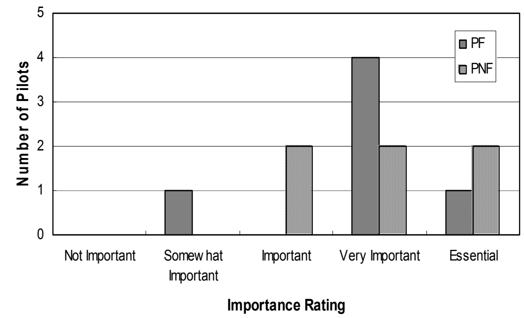

Figure 8. Pilot ratings of the importance of adding realistic ATC communications simulation to pilot simulator training. (n=12 pilots).

Procedural Errors

Analyses of the flight simulator test session videotapes were conducted to determine whether the presence or absence of frequency chatter affected procedural errors either in air-ground communications (e.g., readback errors) or in instrument flying (e.g., failure to identify a navigation aid). Overall, very few errors of either type were found in either the FC or No-FC conditions in this study and the average number of errors did not differ reliably between the two conditions. The provision of high levels of frequency chatter simulation in this study did not appear to affect the aircrew's execution of routine communications and instrument procedures in this study.

Post-test Questionnaire

A post-test questionnaire was completed by all pilots individually following the completion of the flight simulation test session. The first question asked each pilot to rate the importance of adding ATC communication simulation to pilot simulator training. Figure 8 shows the frequency distribution of ratings given by the pilots to each importance rating category. Nine of the twelve pilots (75%) rated the addition of realistic ATC communications simulation as very important or essential. None of the pilots rated the addition of ATC communications simulation to pilot simulator training as not important.

The questionnaire also required pilots to evaluate the relative importance to pilot simulator training of the two keys elements of realistic ATC communications: ownship communications and frequency chatter. Of the 12 pilots responding, ten (83%) indicated that *both* elements were import in simulator training. One of the pilots (8%) regarded ownship communications as solely import and one regarded frequency chatter as the important element.

A third item on the questionnaire asked participating pilots when, in the instrument and advanced flight training regimen of pilots, realistic ATC communications simulation is needed. Four of the twelve pilots (33%) indicated that it should be provided throughout basic instrument flight training. The remaining pilots (67%) thought that realistic ATC communications simulation should be provided only at the end of instrument training or whenever the trainee has achieved instrument flight skill proficiency. None of the pilots interviewed thought that it should be provided only in instrument proficiency evaluations or only during simulations used for awarding advanced ratings or licenses.

General Discussion

In the two earlier studies assessing the impact of high fidelity ATC communications simulation, both crew communications and crew workload were significantly effected. These studies examined the combined effects of two elements of realistic ATC communications simulation, ownship communications and frequency chatter, in line-oriented simulation scenarios. The present study examined the individual contribution of one of these elements, frequency chatter, on crew performance and workload while providing realistic ownship communications throughout the test scenario.

The study found clear evidence of the impact of frequency chatter simulation on crew mental workload as evidenced by the significant increases in the NASA TLX workload ratings by crews in the FC condition. The ratings indicate that the workload impact of frequency chatter is on mental demand and effort of both crewmembers. There was no indication of a larger increase in rated workload for the PNFs when compared to PFs, despite the fact that the PNF was the crewmember primarily responsible for all ATC communications. While workload ratings showed a significant effect of frequency chatter simulation, there was no evidence that it affected the crew in its execution of instrument and communications procedures. The lack of an impact on crew performance in this study is in agreement with the findings of the previous study evaluating ATC communications simulation in airline training environment. The increases in workload ratings for the FC condition in the present study suggest, however, that a more demanding task environment may well have led to more crew procedural errors. Crew comments collected following the test session supports this hypothesis (see Appendix A).

The task of monitoring radio frequencies is considered secondary to those of aircraft control and navigation. The role that simulation of frequency chatter in flight simulator training may play is to increase secondary task loading beyond that periodically provided by the occasional ownship communication. The main benefit of FC simulation may be to allow for a systematic increase in crew mental workload throughout the flight scenario. This is possible since FC can be provided on every frequency used in the scenario (e.g., clearance delivery, ground, tower, departure, etc.). FC transmission rates can also be systematically varied from one scenario to another. Both attributes of FC might allow the I/E to increase crew mental workload systematically[10]. The increased workload would also be more realistic than the current strategy of adding equipment malfunctions or weather-related problems.

While the present study did not assess the role of realistic ownship communications, Studies 1 and 2 both showed increased use of ATC as a resource when realistic ownship communication simulation was compared to that provided by I/Es. While realistic ATC ownship communication simulation may have some effect on crew workload, it is likely that the provision of realistic ownship communications simulation elicits resource management behavior from crews comparable to that which they would exhibit in real-world operations. A potential benefit of realistic ATC ownship communications simulation, therefore, may be the ability of I/Es to more accurately assess crew use of ATC as a resource as well as reinforcing crew air-ground communications procedures (e.g., accurate readbacks).

The analyses above suggest a potential for the two primary elements of ATC communications simulation to affect crew behavior differently. Realistic simulation of frequency chatter may serve to increase crew mental workload as crew members need to devote more cognitive resources to the frequency monitoring task. Whether this increased mental workload affects crew performance will depend upon the workload demand of the particular flight simulation scenario. Realistic simulation of ownship communications, in contrast, is more likely to elicit greater crew use of, and interaction with, ATC and therefore may provide I/Es with a more valid assessment of crew behavior than would otherwise be possible. The effective use of ATC resources is an important element of crew performance. Realistic ownship communications simulation appears to be a means by which this aspect of crew resource management behavior may be assessed.

<u>References</u>

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

2. Lee, A.T. (1999). *Advanced technology applications in flight simulator training and evaluation: Voice generation and recognition.* Tech Rpt. No. BRI-TR-121099. Los Gatos, CA: Beta Research, Inc.

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

[10] The effect of increasing FC transmission rate on workload is supported only by anecdotal evidence at this time.

4. Sumwalt, R, Morrison, R, Watson, A. and Taube, E. (1997). What ASRS data tell about inadequate flight crew monitoring. In *Proceedings of the Ninth International Symposium Aviation Psychology*, pp. 977-983.

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

6. Hart, S. G. and Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-1830. Amsterdam: North Holland.